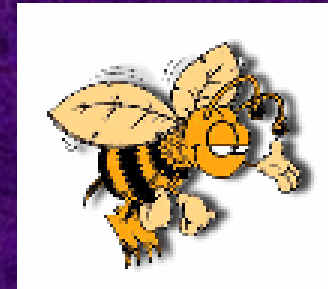


# Introduction to mAdb

Esther Asaki (*in absentia*), John Powell, & John Greene, Ph.D.

- I. Introduction & Overview of the mAdb system
- II. Managing Projects & Uploading Arrays
- III. Initial array quality analysis
- IV. Creating datasets and additional filtering
- V. Basic data analyses and dataset management



September 28, 2004

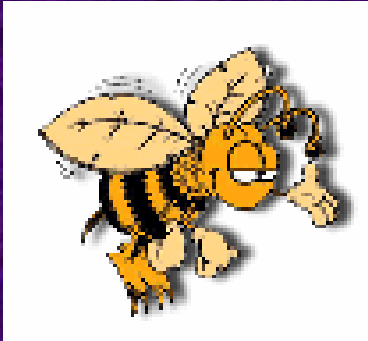
# Logging into the Training Server

- Point your browser at <http://madb-training.cit.nih.gov> – for use in class only!
- Your username is on the card on your desk
- Today's Password is on whiteboard near door
- Don't request a mAdb account on the training server!! – request at [madb.nci.nih.gov](http://madb.nci.nih.gov) or [madb.niaid.nih.gov](http://madb.niaid.nih.gov)
- Do not maximize your browser; leave room to see and click on other windows

# I. Introduction & Overview of the mAdb system

# *mAdb* BioInformatics Project

## Goal:



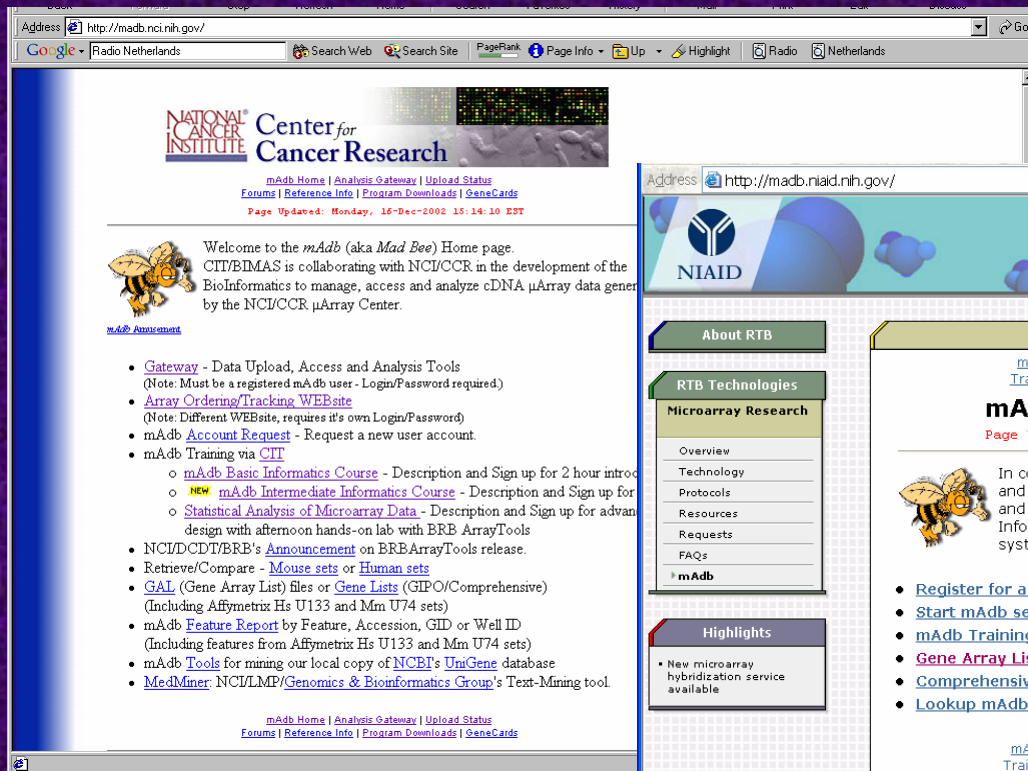
- Provide an integrated set of web-based analysis tools and a data management system for storing and analyzing cDNA/oligo/Affy Gene Expression data using open systems design, focusing on 2 color array slides.
- System currently supports spotted arrays routinely produced by the NCI, NIAID, and FDA Microarray Centers
- Currently support Axon GenePix, Perkin-Elmer QuantArray, and Arraysuite II / IP Lab image analysis software (Yidong Chen, NHGRI) for two-color, “Pat Brown-type” spotted arrays
- Affymetrix now available after a consultation to learn needed parameters – limited number of chips supported right now (mouse, human, rat)



# mAdb Home Page URLs

<http://madb.nci.nih.gov>

<http://madb.niaid.nih.gov>



The screenshot shows the mAdB Home Page for the National Cancer Institute. The header includes the NCI logo and the text "Center for Cancer Research". Below the header, there is a welcome message from CIT/BIMAS, stating they are collaborating with NCI/CCR in the development of the Bioinformatics to manage, access and analyze cDNA μArray data generated by the NCI/CCR μArray Center. A list of links is provided, including Gateway, Array Ordering/Tracking, Account Request, and mAdB Training. A sidebar on the right contains a table of contents for the mAdB database, including Overview, Technology, Protocols, Resources, Requests, FAQs, and mAdB. The footer includes links to the mAdB Home Page, Analysis Gateway, Upload Status, Forums, Reference Info, Program Downloads, and GeneCards.

Address: <http://madb.nci.nih.gov/>

Google | Radio Netherlands | Search Web | Search Site | PageRank | Page Info | Up | Highlight | Radio | Netherlands

**NATIONAL CANCER INSTITUTE** Center for Cancer Research

[mAdB Home](#) | [Analysis Gateway](#) | [Upload Status](#) | [Forums](#) | [Reference Info](#) | [Program Downloads](#) | [GeneCards](#)

Page Updated: Monday, 16-Dec-2002 15:14:10 EST

Welcome to the *mAdB* (aka *Mad Bee*) Home page. CIT/BIMAS is collaborating with NCI/CCR in the development of the Bioinformatics to manage, access and analyze cDNA μArray data generated by the NCI/CCR μArray Center.

[mAdB Announcement](#)

- [Gateway](#) - Data Upload, Access and Analysis Tools (Note: Must be a registered mAdB user - Login/Password required)
- [Array Ordering/Tracking WEBSITE](#) (Note: Different WEBSITE, requires it's own Login/Password)
- mAdB [Account Request](#) - Request a new user account.
- mAdB Training via [CIT](#)
  - [mAdB Basic Informatics Course](#) - Description and Sign up for 2 hour intro
  - [NEW mAdB Intermediate Informatics Course](#) - Description and Sign up for
  - [Statistical Analysis of Microarray Data](#) - Description and Sign up for advanced design with afternoon hands-on lab with BRB ArrayTools
- NCI/DCD/BRB's [Announcement](#) on BRBArrayTools release.
- Retrieve/Compare - [Mouse sets](#) or [Human sets](#)
- [GAL](#) (Gene Array List) files or [Gene Lists](#) (GIP/O/Comprehensive) (Including Affymetrix Hs U133 and Mm U74 sets)
- mAdB [Feature Report](#) by Feature, Accession, GID or Well ID (Including features from Affymetrix Hs U133 and Mm U74 sets)
- mAdB [Tools](#) for mining our local copy of NCBT's [UniGene](#) database
- [MedMiner](#) NCI/LMP/Genomics & Bioinformatics Group's Text-Mining tool.

[mAdB Home](#) | [Analysis Gateway](#) | [Upload Status](#) | [Forums](#) | [Reference Info](#) | [Program Downloads](#) | [GeneCards](#)



The screenshot shows the mAdB Home Page for the National Institute of Allergy and Infectious Diseases (NIAID). The header includes the NIAID logo and the text "RESEARCH TECHNOLOGIES BRANCH". Below the header, there is a welcome message from the Microarray Research Facility at NIAID, stating they are collaborating with the Advanced Technology Center at NCI, the Bioinformatics and Molecular Analysis Section (BIMAS), NIH Center for Information Technology offers the mAdB microarray data analysis system. A list of links is provided, including Register for a mAdB Account, Start mAdB session, mAdB Training/Reference Information, Gene Array List, Comprehensive Gene Lists, and Lookup mAdB Features. A sidebar on the left contains a table of contents for the mAdB database, including Overview, Technology, Protocols, Resources, Requests, FAQs, and mAdB. The footer includes links to the mAdB Home Page, Analysis Gateway, Upload Status, Training/Reference, Program Downloads, and GeneCards.

Address: <http://madb.niaid.nih.gov/>

**NIAID** RESEARCH TECHNOLOGIES BRANCH

**Microarray Research**

[mAdB Home Page](#) | [mAdB Gateway](#) | [Upload Status](#) | [Training/Reference](#) | [Program Downloads](#) | [GeneCards](#)

**mAdB** (MicroArray DataBase, a.k.a "mad bee")

Page Updated: Wednesday, 21-Jul-2004 17:08:24 EDT

In collaboration with the Microarray Research Facility at NIAID and the Advanced Technology Center at NCI, the Bioinformatics and Molecular Analysis Section (BIMAS), NIH Center for Information Technology offers the mAdB microarray data analysis system.

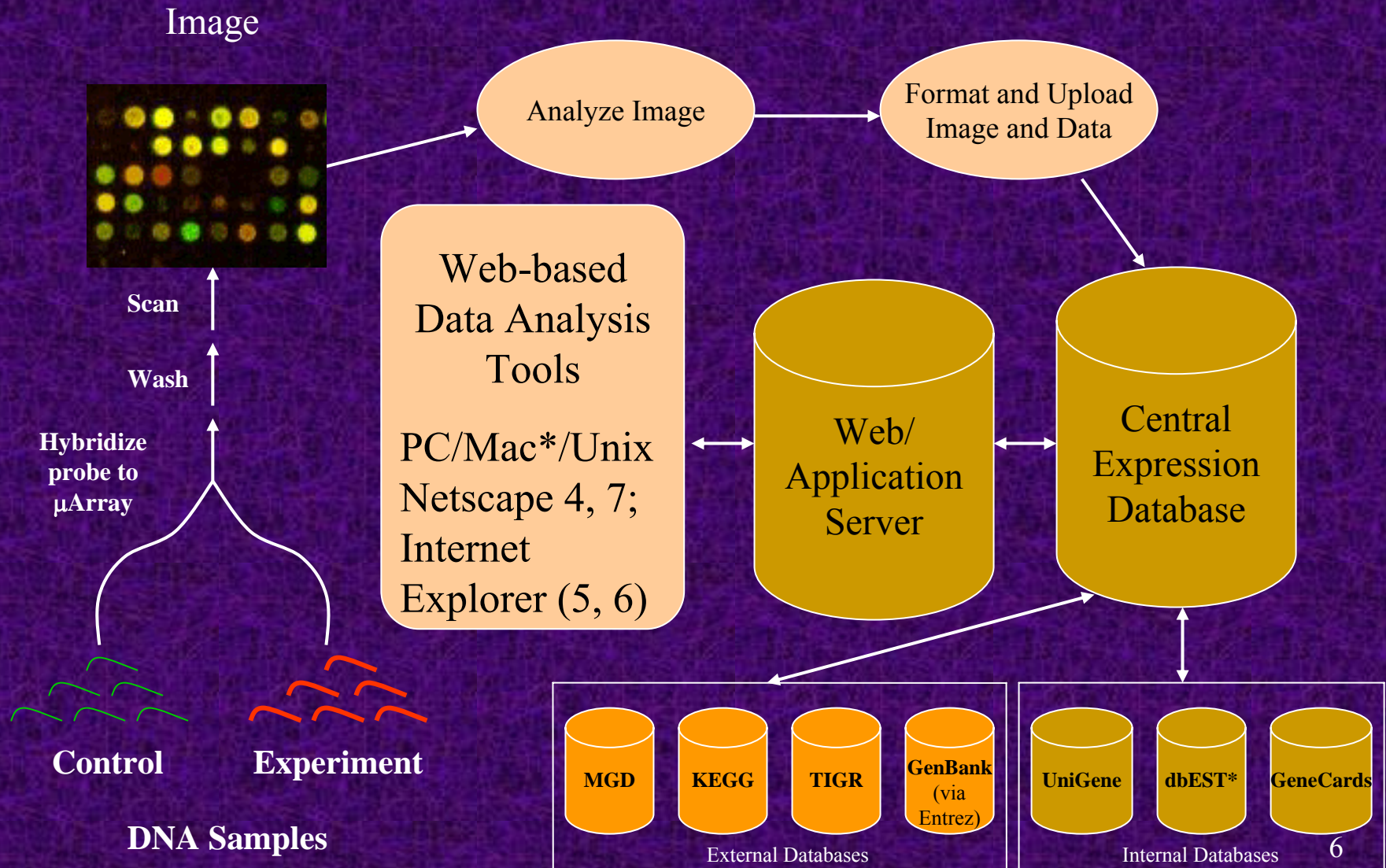
- [Register for a mAdB Account](#)
- [Start mAdB session \(requires mAdB account\)](#)
- [mAdB Training/Reference Information](#)
- [Gene Array List \("GAL"\) files for NIAID MRF Arrays](#)
- [Comprehensive Gene Lists](#)
- [Lookup mAdB Features](#)

[mAdB Home Page](#) | [mAdB Gateway](#) | [Upload Status](#) | [Training/Reference](#) | [Program Downloads](#) | [GeneCards](#)

**NIH Bioinformatics support provided by BIMAS/CBEI/CIT.** We can be contacted by [email](mailto:madb_support@bimas.cit.nih.gov).

For support, please e-mail: [madb\\_support@bimas.cit.nih.gov](mailto:madb_support@bimas.cit.nih.gov)

# Architecture for $\mu$ Array Informatics



# Current *mAdb* Statistics

- 42,222 Arrays uploaded since Feb. 2000 – now average ~1020 per month uploaded over last year
- Roughly  $\frac{3}{4}$  billion cDNA expression measurement points
- 1,180 registered users (NIH and collaborators)
- Among the largest collections of microarray data in the world, although data sharing is determined by each investigator – no one has access to all the data
- MIAME compliance coming very soon
- Can help you get data into public repositories – GEO (NCBI), ArrayExpress (EBI)





# mAdb System Features

- Gene Discovery
  - Outlier detection – row retrieval tools
  - Scatter plots
  - Ad hoc keyword queries
  - Multiple array viewer
- Class Comparison
  - t-test; Wilcoxon; ANOVA; Kruskal-Wallis
- Class Prediction
  - PAM classifier
- Class Discovery (unsupervised)
  - Clustering – Hierarchical, K-means, SOMs
  - Multidimensional Scaling
  - Principal Components Analysis
- Pathway summary – GO, KEGG, BioCarta
- Boolean comparison of data

**Class #412 -  
Analyzing  
Microarray  
Data using  
the mAdb  
System**





# Live Demo

# Home Page Notes

- Special Notices to Users
- Analysis Gateway link
- Account Requests link
- Array Tracker link – N.B. separate login & password!
- Training signup links
- GAL/GIPO links

# mAdb GAL files

## Current (2002) NCI Production Gene Array List Files (GAL files) ( blocks x columns x rows)

- **NEW** [Earlier NCI production printings](#)
- [Custom printings](#)
- [NLAIID printings](#)
- **NEW** [FDA printings](#)
- [Mini-lymphochip GAL files](#) (restricted to registered users)

Human Array Sets			
GAL File	Array Sets		
<a href="#">Hs-UniGEM2-v2px-32Bx18Cx18R.gal</a> Generated Tuesday, 21-May-2002 09:21:59 EDT Note: Also use for 2.1px, 2.3px, 2.4px, 2.5px, 2.6px, 5.0px See below for special 3.5px gal file <b>NEW</b> See below for special 4.0px gal file <b>NEW</b> See below for special 4.1px gal file <b>NEW</b> See below for special 4.2px gal file	Hs-UniGEM2-v2.4p1 Hs-UniGEM2-v2.4p4 Hs-UniGEM2-v2.4p9 Hs-UniGEM2-v2.5p3 Hs-UniGEM2-v2.5p6 Hs-UniGEM2-v2.6p2 Hs-UniGEM2-v2.6p7 Hs-UniGEM2-v2.6p10 Hs-UniGEM2-v5.0p1 Hs-UniGEM2-v5.0p4 Hs-UniGEM2-v5.0p7 Hs-UniGEM2-v5.0p10 Hs-UniGEM2-v5.0p14 Hs-UniGEM2-v5.0p17	Hs-UniGEM2-v2.4p2 Hs-UniGEM2-v2.4p5 Hs-UniGEM2-v2.5p4 Hs-UniGEM2-v2.5p7 Hs-UniGEM2-v2.6p3 Hs-UniGEM2-v2.6p8 Hs-UniGEM2-v5.0p2 Hs-UniGEM2-v5.0p5 Hs-UniGEM2-v5.0p8 Hs-UniGEM2-v5.0p12 Hs-UniGEM2-v5.0p15 Hs-UniGEM2-v5.0p18	Hs-UniGEM2-v2.4p3 Hs-UniGEM2-v2.4p8 Hs-UniGEM2-v2.5p5 Hs-UniGEM2-v2.5p11 Hs-UniGEM2-v2.6p6 Hs-UniGEM2-v2.6p9 Hs-UniGEM2-v5.0p3 Hs-UniGEM2-v5.0p6 Hs-UniGEM2-v5.0p9 Hs-UniGEM2-v5.0p13 Hs-UniGEM2-v5.0p16
<a href="#">Hs-UniGEM2-v3.5px-32Bx19x17R.gal</a> Generated Tuesday, 21-May-2002 09:33:10 EDT	Hs-UniGEM2-v3.5p1 Hs-UniGEM2-v3.5p2		
<a href="#">Hs-UniGEM2-4.0px-32Bx18Cx18R.gal</a> Generated Monday, 25-Nov-2002 15:03:35 EST	Hs-UniGEM2-v4.0p2 Hs-UniGEM2-v4.0p6 Hs-UniGEM2-v4.0p9	Hs-UniGEM2-v4.0p4 Hs-UniGEM2-v4.0p7 Hs-UniGEM2-v4.0p10	Hs-UniGEM2-v4.0p5 Hs-UniGEM2-v4.0p8 Hs-UniGEM2-v4.0p11
<a href="#">Hs-UniGEM2-4.1px-32Bx18Cx18R.gal</a> Generated Monday, 25-Nov-2002 15:34:59 EST	Hs-UniGEM2-v4.1p1		

• Shows the actual GAL (Gene Array list) files – link block, row, column to what DNA is spotted there

• One printset layout is usually used for many lots of slides

• Please e-mail mAdb support if you cannot find your GAL file listed

# Application Program Downloads

## mAdb Program Downloads

Page Updated: Friday, 15-Aug-2003 08:45:58 EDT

	Program	Description	Author	Version	Updated	Download	Manual
<b>Axon Inc. Software</b> This is commercial, licensed software and the GenePix application requires a "dongle" attached to the parallel port to run. The manual is accessible to all. <a href="#">Axon's Web Site</a>	 <b>GenePix Pro 5</b>	Fully integrated acquisition and analysis software for the GenePix 4000, 4100 & 4200. Download to a folder of your choice and then run to start the installation process.		5.0.1.13 <a href="#">History</a>	8/15/2003 (Posted here 8/15/2003)	<a href="#">Download</a>	<a href="#">Users Guide &amp; Tutorial</a> (PDF)
<b>Axon Inc. Software</b> This is commercial, licensed software and the GenePix application requires a "dongle" attached to the parallel port to run. The manual is accessible to all. <a href="#">Axon's Web Site</a>	 <b>GenePix Pro 4</b>	Fully integrated acquisition and analysis software for the GenePix 4000 & 4100. Download to a folder of your choice and then run to start the installation process.		4.0.1.17 <a href="#">History</a>	(Posted here 3/12/2003)	<a href="#">Download</a>	<a href="#">Manual</a> <a href="#">Axon Scanner Manual</a> (PDFs)
<b>Axon Inc. Software</b> This is commercial, licensed software and the GenePix application requires a "dongle" attached to the parallel port to run. The manual is accessible to all. <a href="#">Axon's Web Site</a>	 <b>GenePix Pro 3</b>	Fully integrated acquisition and analysis software for the GenePix 4000A. Download to a folder of your choice and then run to extract the installation files. Then run the extracted file setup.exe and follow installation instructions		3.0.6.89 <a href="#">History</a>	(Posted here 02/18/2002)	<a href="#">Download</a>	<a href="#">Manual</a> <a href="#">Axon Scanner Manual</a> (PDFs)
<b>Stanford Genome Analysis Group Software</b> It is available free of charge to academic and non-profit institutions. <a href="#">Eisen Lab Download Site</a>	 <b>ScanAlyze</b>	Image Analysis (extracts data from fluorescence images of arrays)	<a href="#">Michael Eisen</a>	2.44	11/15/99	<a href="#">Download</a>	<a href="#">Manual</a> (PDF)
	 <b>Cluster</b>	Perform Hierarchical Clustering, Self-organizing Maps, k-Means Clustering, and More	<a href="#">Michael Eisen</a>	2.11.01	7/10/2000 (Posted here 10/26/2000)	<a href="#">Download</a>	<a href="#">Manual</a> (PDF)
	 <b>Tree View</b>	Graphical Viewing and Browsing of Cluster Results	<a href="#">Michael Eisen</a>	1.5	04/2000 (Posted here 2/8/2001)	<a href="#">Download</a>	
<b>EASE: Expression Analysis Systematic Explorer</b> Developed by the Laboratory of Immunopathogenesis and Bioinformatics, SAIC Frederick <a href="#">EASE Web Site</a>	 <b>EASE</b>	For finding "biological meaning" of gene lists via three functions: biological theme over-representation analysis, creation of annotation tables, and automated loading of genes into various online tools.	<a href="#">Doug Hosack</a>	<a href="#">Revision history</a>	<a href="#">Current version</a>	<a href="#">Link to Download</a>	<a href="#">Online help</a> (Online)
<b>MAExplorer</b> Developed by and Available from LECB/FCRF/NCI. <a href="#">MAExplorer Web Site</a>	 <b>MAExplorer</b>	A Java data mining application for gene expression data using a variety of statistical, clustering, direct-manipulation graphical, spreadsheet and Web access methods.	<a href="#">Peter Lemkin</a>	<a href="#">Revision History</a>	<a href="#">Current version</a>	<a href="#">Link to Download</a>	<a href="#">Manual</a> (Online) <a href="#">Use with mAdb data</a> (PDF)

GenePix 5 is now supported, with important bug fix for spot images

Page accessible from NIH network only

Prefer GenePix updates obtained from this page – validated to work with mAdb




# Reference Page

## Reference Information

---

Page Updated: Thursday, 27-May-2004 16:52:14 EDT

- mAdb Reference Documents
  - Introduction to mAdb (CIT class #972) Training Slides with Labs: [PowerPoint](#) or [PDF](#)  
Updated Tuesday, 01-Jun-2004 12:27:04 EDT
  - Analyzing Microarray Data with the mAdb System (CIT class #974) Training Slides
    - Lecture Slides: [PowerPoint](#) or [PDF](#)  
Updated Monday, 10-May-2004 17:45:09 EDT
    - Hands-on Labs: [PowerPoint](#) or [PDF](#)  
Updated Monday, 10-May-2004 17:45:24 EDT
  -  Uploading Affymetrix Data to mAdb: [PDF](#)  
Updated Thursday, 27-May-2004 16:04:17 EDT  
NOTE: You must request permission from [mAdb support](#) before uploading Affymetrix Data.
  - Increasing Upload Speed with Internet Explorer on the PC: [Word](#) or [PDF](#)  
Updated Wednesday, 14-May-2003 10:23:06 EDT

- Also links to protocols
- GenePix manuals
-  N.B. Still must request Affy privileges be turned on for your account

## II. Managing Projects and Uploading Arrays

# Data Upload

- Login to Analysis Gateway page
  - change password if first-time user (case sensitive)
- Create project - logical organization for arrays
- Grant project access to others (if desired)
- Return to gateway and select project
- Select **type** of array for project
  - Spotted OR
  - Affymetrix (need to request permission via e-mail for first usage so we can give you needed parameters)

# mAdb Gateway- link for Project Creation & Management

## *mAdb Gateway*

**NEW** Create/Manage Projects link under Management Tools below. From there you can Create, Edit and Delete (empty projects) projects as well as Manage Access to projects.

Choose one or more Projects, select a Tool and Continue  
or access previously extracted data located in **ncidemo**'s:  
[Temporary](#) or [Permanent](#) area

**Projects:**

XX guest - Time Course Demo Set #1  
XX guest - Time Course Demo Set #2  
XX guest - Repeats and Reciprocal Retests Demo Set #3  
XX guest - Multiple Types Demo Set #4  
AU ncidemo - my project  
AU ncidemo - Oligo and cDNA

Note: Tools marked with "XX" only support selection of one project

**Tool:**

Project Summaries Report

Continue

### Uploading Links

- [Upload](#) Array data
- [Status](#) of Uploads
- [Upload](#) Identifier lists
- [Manage](#) Identifier lists



### Management Tools

- [Create/Manage](#) Projects
- [Manage](#) User Profile



[Access](#) Training/Public Datasets

[Access](#) Additional Public Datasets



# User Profile Management

## Managing User Profile

---

[Change](#) Your Password

[Update](#) Your User Profile

Profile for "ncidemo" last modified on Sep 03, 2004 at 15:03:08

**Title** Mr.

**First Name** DEMO

**Middle Initial**

**Last Name** NCI

**E-mail** jip@helix.nih.gov

**Position**

**Affiliation**

**NIH Address** 12A/2033 Bethesda, MD 20892

**Work Phone**

**Fax**

You have chosen to NOT Subscribe to the E-Newsletter

# Managing Projects

## Managing Projects

---

### Create New Project

Shown below are existing Projects for which "ncidemo" is an administrator.  
Projects are ordered first by the Creator and then by the Creation Date  
In the Access List, **Bold** indicates a user with administrative access

### Management Options

mAdb ID# 160 created by "ncidemo" on Jun 26, 2000 at 15:47:00 contains 10 Arrays

**Project Title:** my project

**Description:** Description by jip. Altered @1:00pm on 8/31/2004 and altered again by "easaki" on 9/1/2004

**Comments:** Comments by jip. Altered 8/31/2004

**Access List:** easaki, **jmgreene**, jpowell, **ncidemo**

### Management Options

mAdb ID# 1195 created by "ncidemo" on May 30, 2002 at 13:53:50 contains 10 Arrays

**Project Title:** Oligo and cDNA

**Description:** mixture of oligo and cDNA arrays

**Comments:** for IM class

**Access List:** easaki, **ncidemo**

### Management Options

mAdb ID# 2874 created by "ncidemo" on Jun 24, 2004 at 13:27:42 contains no arrays

**Project Title:** Drug abcd

**Description:** jkas,ldkjflk

**Comments:** ;lkjljk,alsdjfsklk

**Access List:** easaki, **ncidemo**

# Create New Project

**Create New Project**

---

created by ncidemo

**Project Title:**

**Description:**

**Comments:**

- **A Project is a logical grouping of your arrays**

# Project Management Options

## Project Management Options

---

mAdb ID# 160 created by "ncidemo" on Jun 26, 2000 at 15:47:00 contains 10 Arrays

**Project Title:** my project

**Description:** Description by jip. Altered @1:00pm on 8/31/2004 and altered again by "easaki" on 9/1/2004

**Comments:** Comments by jip. Altered 8/31/2004

**Access List:** **easaki, jmgreene, jpowell, ncidemo**

**Click**    **Options available for this Project**



Can not be deleted - contains 10 Arrays

[Edit](#)    To modify the Project Information (Title, Description, Comments)

[Add](#)    To Add user(s) to the Access List for this Project

[Remove](#)    To Remove user(s) from the Access List for this Project

[Privileges](#)    To Grant or Revoke User(s) Administrative/Upload privileges for this Project

[Return](#) to Managing Projects

**Bold names on access list indicate administrative privileges for account**



# Project Access

## Add User(s)

mAdb ID# 160 created by "ncidemo" on Jun 26, 2000 at 15:47:00 contains 10 Arrays

**Project Title:** my project

**Description:** Description by jip. Altered @1:00pm on 8/31/2004 and altered again by "easaki" on 9/1/2004

**Comments:** Comments by jip. Altered 8/31/2004

**Access List:** easaki, jimgreene, jpowell, ncidemo

The List below includes **ALL mAdb users** not already having access to this project.

Add User(s)

Reset Form

Cancel

### Check to select User(s) to add to this project

▼ Last name, First name ( Login )

- ☐ Abdool, Karen ( abdoolk )
- ☐ Abul-Hassan, Khaled ( hassank )
- ☐ Ajay, Dr ( ajay\_dr )
- ☐ Akagi, Keiko ( akagik )
- ☐ Aksamit, Robert ( aksamit )
- ☐ Al-Timimi, Ali ( altimima )
- ☐ Albert, Paul ( albertp )
- ☐ Aleman, Claudina ( alemanc )
- ☐ Alexander, H. Richard ( ralexander )
- ☐ Alizadeh, Ash ( alizadeh )
- ☐ Alkharouf, Nawal ( nalkhar )
- ☐ Amornphimoltham, Panomwat ( pa79w )
- ☐ Amundson, Sally ( amundson )
- ☐ Anderson, Soni ( andersso )
- ☐ Andersson, John ( jandersson )
- ☐ Andreola, Fausto ( andreolf )

▼ Last name, First name ( Login )

- ☐ Mazzanti, Chiara ( chiara )
- ☒ McCarty, Tom ( tmccarty )
- ☐ McConnell, Melanie ( melanie.mccconnell )
- ☐ McDonald, Shannon ( slmcdonald )
- ☐ McKee, Marian ( mmckee )
- ☐ McNeil, Nicole ( mcneiln )
- ☐ McNeill, Megan ( mmcneill )
- ☐ McShane, Lisa ( mcshanel )
- ☐ Medjahed, Djamel ( medjahed )
- ☐ Mejido, Josef ( mejido )
- ☐ Melani, Raffaella ( rmelani )
- ☐ Meletiadis, Joseph ( meletiaj )
- ☐ Melillo, Giovanni ( melillo )
- ☐ Meltzer, Stephen ( umddemo )
- ☐ Memon, Sarfraz ( memonsa )
- ☐ Menard, Cynthia ( menardc )

Adding a user allows that mAdb account holder to view your arrays in a project and work with the data to create filtered datasets

# User Privileges

## Change User(s) Privileges

mAdb ID# 160 created by "ncidemo" on Jun 26, 2000 at 15:47:00 contains 10 Arrays

**Project Title:** my project

**Description:** Description by jip. Altered @1:00pm on 8/31/2004 and altered again by "easaki" on 9/1/2004

**Comments:** Comments by jip. Altered 8/31/2004

Check/UnCheck as appropriate to select privileges

Admin Upload

▼	▼		Last name, First name ( Login )
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	AU	Asaki, Esther ( easaki )
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	AU	Greene, John ( jmgreene )
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	AU	NCI, DEMO ( ncidemo )
<input type="checkbox"/>	<input type="checkbox"/>	--	Powell, John ( jpowell )

Record Changes

Reset Form

Cancel

Privileges allow user to:

- Upload Arrays
- Administer access to arrays and edit project, array, and dataset descriptions

# mAdb Tool Gateway- link for uploading

## *mAdb* Gateway

**NEW** Create/Manage Projects link under Management Tools below. From there you can Create, Edit and Delete (empty projects) projects as well as Manage Access to projects.

Choose one or more Projects, select a Tool and Continue  
or access previously extracted data located in **ncidemo**'s:

[Temporary](#) or [Permanent](#) area

**Projects:**

XX guest - Time Course Demo Set #1  
XX guest - Time Course Demo Set #2  
XX guest - Repeats and Reciprocal Retests Demo Set #3  
XX guest - Multiple Types Demo Set #4  
AU ncidemo - my project  
AU ncidemo - Oligo and cDNA

Note: Tools marked with "X" only support selection of one project

**Tool:**

Project Summaries Report

Continue

### Uploading Links

- [Upload](#) Array data
- [Status](#) of Uploads
- [Upload](#) Identifier lists
- [Manage](#) Identifier lists



### Management Tools

- [Create/Manage](#) Projects
- [Manage](#) User Profile



[Access](#) Training/Public Datasets

[Access](#) Additional Public Datasets

# Spotted Array Data Upload

- Fill in experimental info for each array
  - Pick Print Set
  - Select image file of array
  - Select data file for array
- Submit and confirm upload
- Check status page to display progress
- Close browser when finished (for security)




# Affymetrix Data Upload


- Select:
  - Data File (Metrics - .txt file)
  - CEL file
- Fill in Experiment data
- Submit and confirm upload
- Check status page to display progress
- Close browser when finished (for security)

# Uploading Spotted Arrays

Project: MYTEST PROJECT

Print Sets printed in the past 365 days 

**Experiment Information**

Prints since **May 08, 2002** (past 180 days). 

Array Print Set: Hs-UniGEM2-v4.0p9-061102

Array Name: Hs-UG4p9 Suggested form: HsOC2p13-45

Short Description: TIMP-4 time course - 1 hour

Long Description: Optional description of this experiment - may include dosage or concentration information, animal or cell line identifiers, etc.

Probe: Untested Control Channel A (generally **Cy3** tagged) Channel B (generally **Cy5** tagged)

Probe Label: Cy3 Cy5

**Composite Image & Arraysuite Sample Intensities or GenePix GPR Files**

Image File: myarray1.jpg

Data File: myarray1.gpr

**NOTICE:** A confirmation step has been added to the upload process. After the information and files are uploaded, you will be presented a screen with information about the files uploaded. Unless you select "Confirm" the Uploaded Files will not be placed on the Queue for insertion into mAdb.

Default shows printings  
from last six months

# Confirming Upload

## NCI/NIH *mAdb* Data Loading Gateway

---

### Upload Confirmation:

Details from a preliminary inspection of the Intensity and Image files are provided below.  
You may Confirm or Cancel the uploading process.

#### Data File:

C:\Documents and Settings\greenej1.NIH\Desktop\DataFile.txt


#### Image File:

C:\Documents and Settings\greenej1.NIH\Desktop\ImageFile.img

Data file appears to be: Axon Text Format (GenePix Pro 3/4 Results)

Number of Data Values appears to be : 8837

Image Format: JPEG



[Return to Data Loading Page](#)

[Return to MicroArray Home Page](#)

[mAdb Home](#) | [Analysis Gateway](#) | [Upload Status](#)  
[Forums](#) | [Reference Info](#) | [Program Downloads](#) | [GeneCards](#)

You should check that the image and file type appear correct and that the file line count is roughly equal to the number of spots on the array

# Adding Affy Arrays

## Upload MAS5 Analysis Data to:my project

Note the  marks the link which lead to detailed help on required Affymetrix file format

### Affymetrix Files for Upload

<b>Data File:</b>	<input type="text"/>	<input type="button" value="Browse..."/>
<b>Cel File:</b>	<input type="text"/>	<input type="button" value="Browse..."/>

- Browse to Metrics (\*.txt) file for the Data File box
- Browse to the corresponding .CEL file in second box



# Adding Affy Arrays

## Confirm Affymetrix Genechip Data

### Experiment Information

You have uploaded Absolute Analysis data for a Human Genome Array U95A genechip.

The Data have not been scaled in your analysis.

Please check/complete the information on this page. Click the Confirm button to complete the upload process or use the Cancel button to abort and start again.

**Uploaded Data File:** C:\GeneChip\TESTDATA\Gene Logic

**Spike\92453hgu95a11\_test.txt**

**Uploaded CEL File:** C:\GeneChip\TESTDATA\Gene Logic

**Spike\92454hgu95a11.cel**

Fields labled with \*\* are mandatoray.

**Array Print Set:** U95A

**Array Name:** \*\* 92453hgu95a11\_test

**Sample Type:**

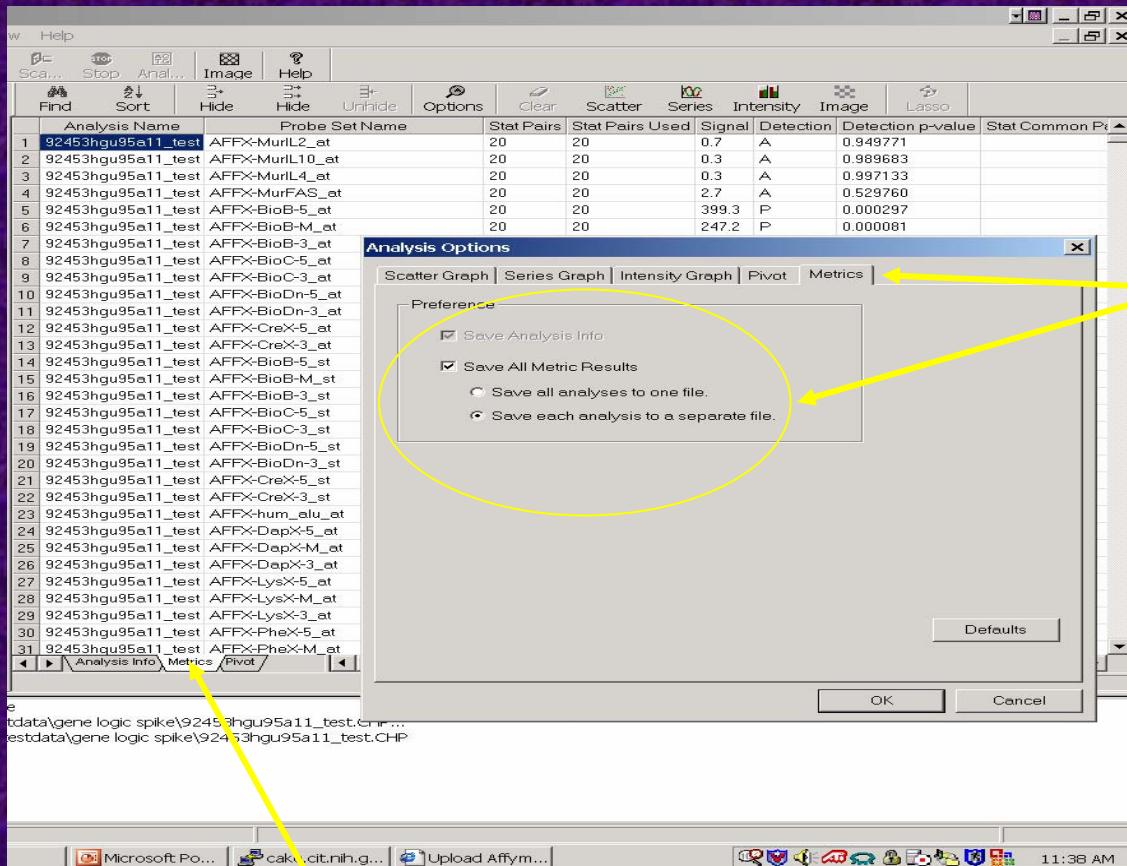
**Sample Description:**

**Comments:**

Confirm

Cancel

# Affymetrix – CHP file

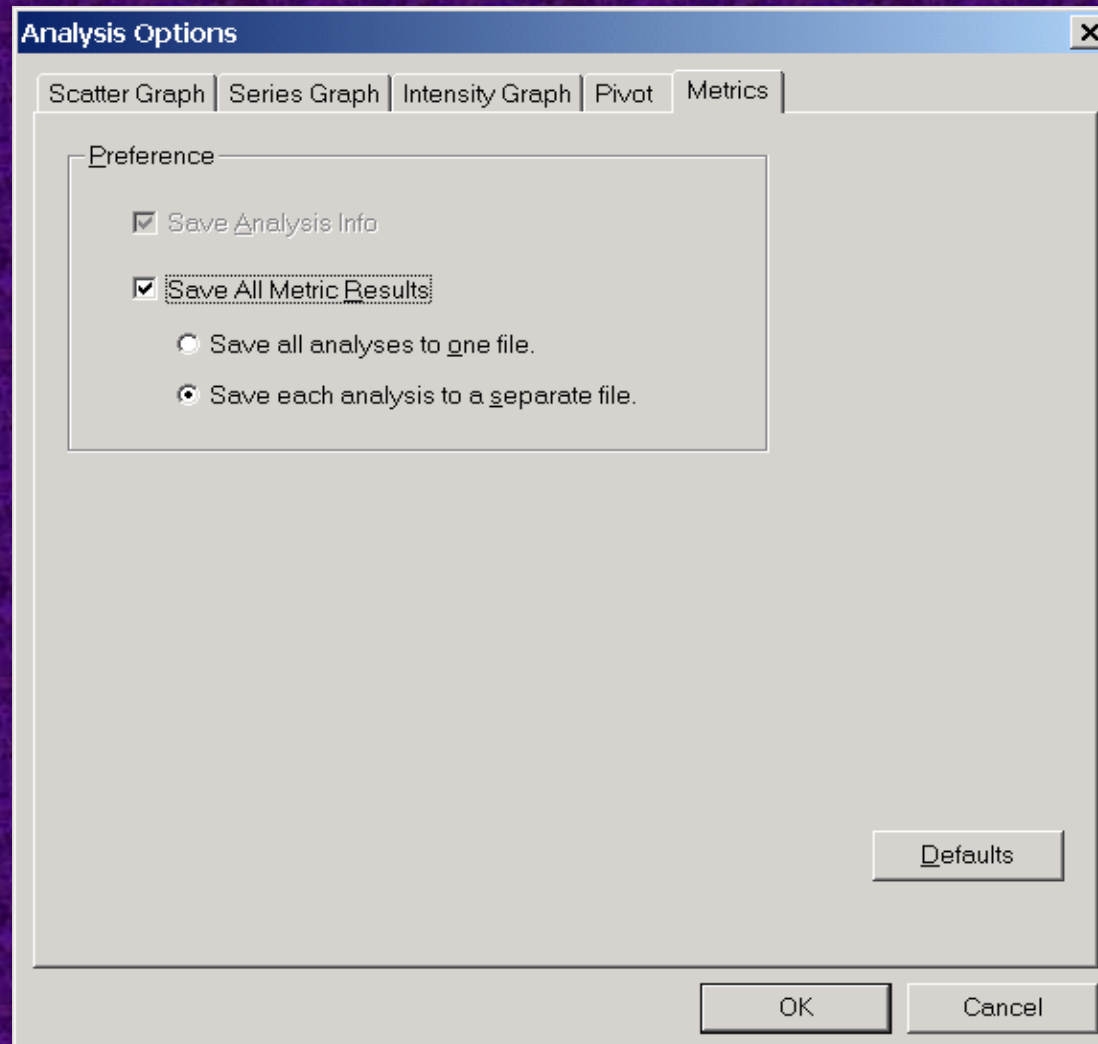


Set Metrics options:

- Save all Metric Results
- Save each analysis to a separate file

Select Metrics tab before saving

# Affymetrix – CHP file Metrics options



# Upload Status

- Shows your arrays and totals for all users
- Two step process:
  - Data is parsed and entered into Sybase db
  - Image is processed and stored
- You can work with data without waiting for image processing to finish

## mAdb WEB Upload Status Report

Status Updated: Tue Sep 28 10:54:29 EDT 2004  
(This page refreshes every 10 minutes)

### Other mAdb WEB Upload Reports:

Graphical summary by [month](#) (past 12 months) or by [day](#) (past 90 days)

Details of arrays [queued](#) for processing

Details of arrays uploaded within the past [24 hours](#), [7 days](#), [30 days](#) or [all](#)

### mAdb login      Arrays      Status

ncidemo      0      Queued for/or loading into mAdb

**Total all Users** 0      Queued for/or loading into mAdb

ncidemo      0      Loaded; Queued for/or Image processing

**Total all Users** 0      Loaded; Queued for/or Image processing

### Activity for the past 30 days

ncidemo      0      Processing completed

**Total all Users** 1037      Processing completed (332 Affymetrix, 705 Spotted)

ncidemo      0      Canceled, UnConfirmed, Bad Files/Rejected Submissions;

**Total all Users** 44      Canceled, UnConfirmed, Bad Files/Rejected Submissions



# GenePix Analysis Notes

- Download correct GAL file from mAdb
- Carefully grid each block
- Allow program to “Find spots” and adjust spot size
- Set option to “Analyze absent spots”
- Adjust JPEG for desired contrast/brightness
- Analyze spots

# Spotted Array Uploading Notes

- Include the slide number scratched on the slide as part of the Array Name, which will act as a unique identifier
- If array print was printed more than 6 months ago, extend time frame using the pulldown menu and press “Show” button
- When uploading to a new project, array prints from all species will be displayed
- When uploading to a project with arrays, only array prints using the same species in the project will be displayed.

# Common Spotted Array Errors

- **Common Upload Errors**
  - Choosing wrong print set
  - Loading GAL file, Excel file, or Set Up file in place of GenePix data (.gpr) file
  - Loading multi-image TIFF file instead of composite, single image JPEG or PICT file
- **Common GenePix Errors**
  - Setting incorrect option for “Analyze Absent Feature” (box should be checked) – results in truncated blocks
  - Deleted blocks
  - Improper gridding

# Affymetrix Analysis Notes

- Run chip through fluidics station to get CEL file
- Analyze CEL file (usually scale all spots to 500)
- With CHP file open, set analysis options on metrics tab as:
  - “Save Analysis Info”
  - “Save each analysis to a separate file”
- Click on Metric tab
- Save file as .... Xxxx.txt
- Note: If uploading comparison data, then upload absolute baseline data first.



# Ability to copy or move arrays between projects

- Need administrative access to both projects
- Create a “trash” project to “delete” unwanted arrays

## mAdb Copy/Move Arrays

Options 

**Move**

Selected Arrays

To Project:

Arrays from

**Project 1038:** Multiple Types Demo Set #4

**Created on:** Mar 5 2002 9:02AM

**Description:** Example of repeats of different types (for example tissue, cell lines, animal strain)

Array Selection 

– A

Submit

	A	mAdbID: Array Name & Short Description
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28733: Mm-Incyte-vlp1-1 Sample 1/Type A
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28742: Mm-Incyte-vlp1-10 Sample 5/Type B
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28734: Mm-Incyte-vlp1-2 Sample 2/Type A
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28735: Mm-Incyte-vlp1-3 Sample 3/Type A
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28736: Mm-Incyte-vlp1-4 Sample 4/Type A
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28737: Mm-Incyte-vlp1-5 Sample 5/Type A
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28738: Mm-Incyte-vlp1-6 Sample 1/Type B
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28739: Mm-Incyte-vlp1-7 Sample 2/Type B
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28740: Mm-Incyte-vlp1-8 Sample 3/Type B
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28741: Mm-Incyte-vlp1-9 Sample 4/Type B

# Ability to re-order arrays within a project

## Order Arrays within Project

Note: This tool changes the order designation for arrays within this project. All users who have access to this project will see this order designation.

**Arrays**

↑

Change  
Array  
order.

↓

Mm-Incyte-v1p1-6 Sample 1/Type B  
Mm-Incyte-v1p1-7 Sample 2/Type B  
Mm-Incyte-v1p1-8 Sample 3/Type B  
Mm-Incyte-v1p1-9 Sample 4/Type B  
Mm-Incyte-v1p1-10 Sample 5/Type B  
Mm-Incyte-v1p1-1 Sample 1/Type A  
Mm-Incyte-v1p1-2 Sample 2/Type A  
Mm-Incyte-v1p1-3 Sample 3/Type A  
Mm-Incyte-v1p1-4 Sample 4/Type A  
Mm-Incyte-v1p1-5 Sample 5/Type A

Submit

Cancel

**Change Array Order** by highlighting an array name and using the change array order up and down arrows.

Click the **Submit** button when finished or the **Cancel** button to return to the Analysis Gateway.

# III. Initial Array Quality Analysis

- Signal definition
- Normalization
- Use of log base 2
- Project Summary Report
- Comprehensive Graphical Quality Report

# mAdb Definitions

- Signal - refers to background corrected values (i.e. Target Intensity - Background Intensity).
- Defaults:
  - MEAN Intensity – MEDIAN background (for GenePix)
  - MEAN Intensity – MEAN background (for ArraySuite)
- Normalization factor – initially calculated so that the median overall ratio (Cy5 Signal/ Cy3 Signal) is adjusted to 1.0 (linear; 0.0, log base 2) for each array. Spots with an extremely low signal are excluded from this calculation.



# Need for Normalization of Ratios

- Unequal incorporation of labels (green Cy3 incorporates better than red Cy5)
- Unequal amounts of samples
- Unequal PMT voltage settings

# Whenever possible, use ratios converted to log base 2

- Why? Because it makes variation of ratios more independent of absolute magnitude
- Evens out highly skewed graphs, giving a more realistic sense of variation – upregulated genes graph from 1 to  $\infty$  ; downregulated genes graph crammed between 0 and 1
- Easier interpretation – negative numbers are downregulated genes; positive numbers are upregulated genes

## mAdb Array Histogram

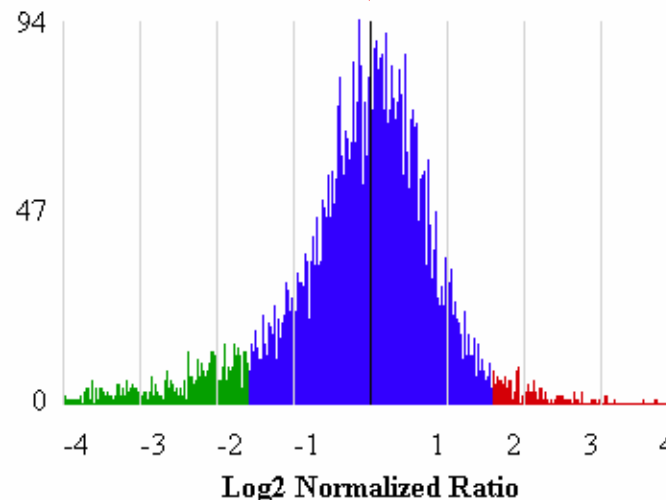
[Comprehensive Graphical Report](#) (Be Patient!)

Array: Mm-Incyte-v1p1-1

Short Description: Sample 1/Type A

Long Description: Add a description

re-centered



Empty wells and flagged spots filtered out

Green: Ratio < 1/3, Red: Ratio > 3

Mean Signal		Median Bkg		Sgl/Bkg		Not Found	Normal. Factor**
Ch A	Ch B	Ch A	Ch B	Ch A	Ch B		
326	455	110	84	3.0	5.4	30%	0.617

Normalization factor is calculated and multiplied by all ratios to re-center array distribution around 1 (linear), equal to 0 in log base 2

# Project Summary

## mAdb Project Summaries 1.0

Retrieve

Array Summaries formatted for MS-Excel





















[Edit](#) Project #1038: Multiple Types Demo Set #4

Created on: Mar 05, 2002

Description: Example of repeats of different types (for example tissue, cell lines, animal strain)

### Summary Statistics

### Array Information

		Mean Signal		Median Bkg		Sgl/Bkg		% Found	Normal Factor	mAdb ID	Uploaded		Array Print	Array	Probe A	Probe B	Short
		Ch A	Ch B	Ch A	Ch B	Ch A	Ch B										
  <a href="#">Edit</a>	1.	326	455	110	84	3.0	5.4	70%	0.626	28733	Mar 5 2002 9:07AM		Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-1	Control	Sample 1	Samp
  <a href="#">Edit</a>	2.	1677	2088	241	160	7.0	13.1	93%	0.769	28742	Mar 5 2002 9:24AM		Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-10	Control	Sample 5/B	Samp
  <a href="#">Edit</a>	3.	880	673	200	364	4.4	1.8	84%	1.055	28734	Mar 5 2002 9:10AM		Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-2	Control	Sample 2	Samp
  <a href="#">Edit</a>	4.	1056	1473	259	154	4.1	9.6	93%	0.658	28735	Mar 5 2002 9:11AM		Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-3	Control	Sample 3	Samp
  <a href="#">Edit</a>	5.	297	493	117	87	2.5	5.7	84%	0.542	28736	Mar 5 2002 9:13AM		Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-4	Control	Sample 4	Samp
  <a href="#">Edit</a>	6.	443	543	123	89	3.6	6.1	83%	0.708	28737	Mar 5 2002 9:15AM		Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-5	Control	Sample 5	Samp
  <a href="#">Edit</a>	7.	499	541	120	101	4.2	5.4	84%	0.858	28738	Mar 5 2002 9:17AM		Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-6	Control	Sample 1/B	Samp
  <a href="#">Edit</a>	8.	626	717	146	113	4.3	6.3	85%	0.890	28739	Mar 5 2002 9:21AM		Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-7	Control	Sample 2/B	Samp
  <a href="#">Edit</a>	9.	1280	1399	272	190	4.7	7.4	93%	0.830	28740	Mar 5 2002 9:22AM		Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-8	Control	Sample 3/B	Samp
  <a href="#">Edit</a>	10.	1113	1371	261	156	4.3	8.8	91%	0.779	28741	Mar 5 2002 9:23AM		Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-9	Control	Sample 4/B	Samp

- Aid to QC – overall array statistics, links to histogram, array image
- If you have admin access to a project, can edit project and array descriptions from “Edit” links here

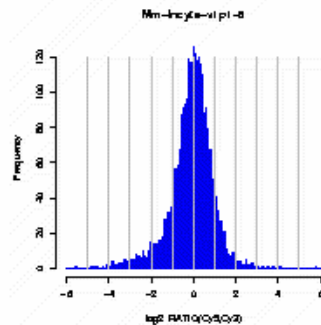


# Comprehensive Graphical Quality Report

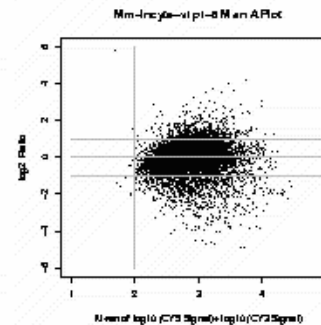
## mAdb Array Report

Signal Calculation: Mean Intensity - Median Background  
Normalization Factor: 0.8174

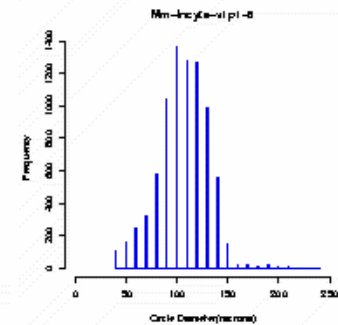
- Accessed from histogram display
- More QC parameters, including:
  - M versus A plot
  - spot size distribution
  - log and linear plots of each channel
  - signal intensity distribution
  - signal/background distribution



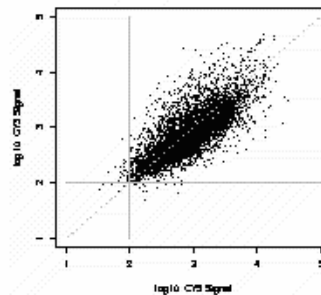
EPS, PDF, PNG



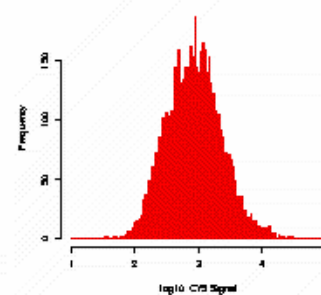
EPS, PDF, PNG



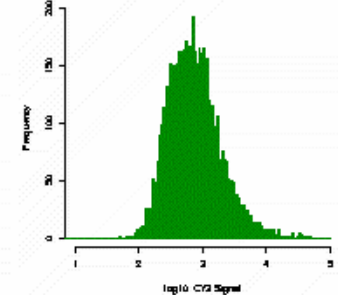
EPS, PDF, PNG



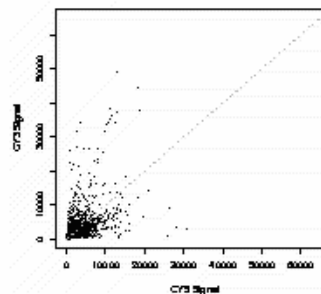
EPS, PDF, PNG



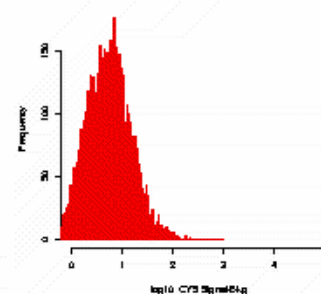
EPS, PDF, PNG



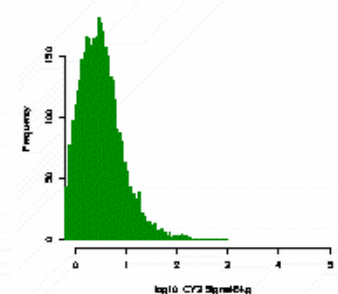
EPS, PDF, PNG



EPS, PDF, PNG



EPS, PDF, PNG

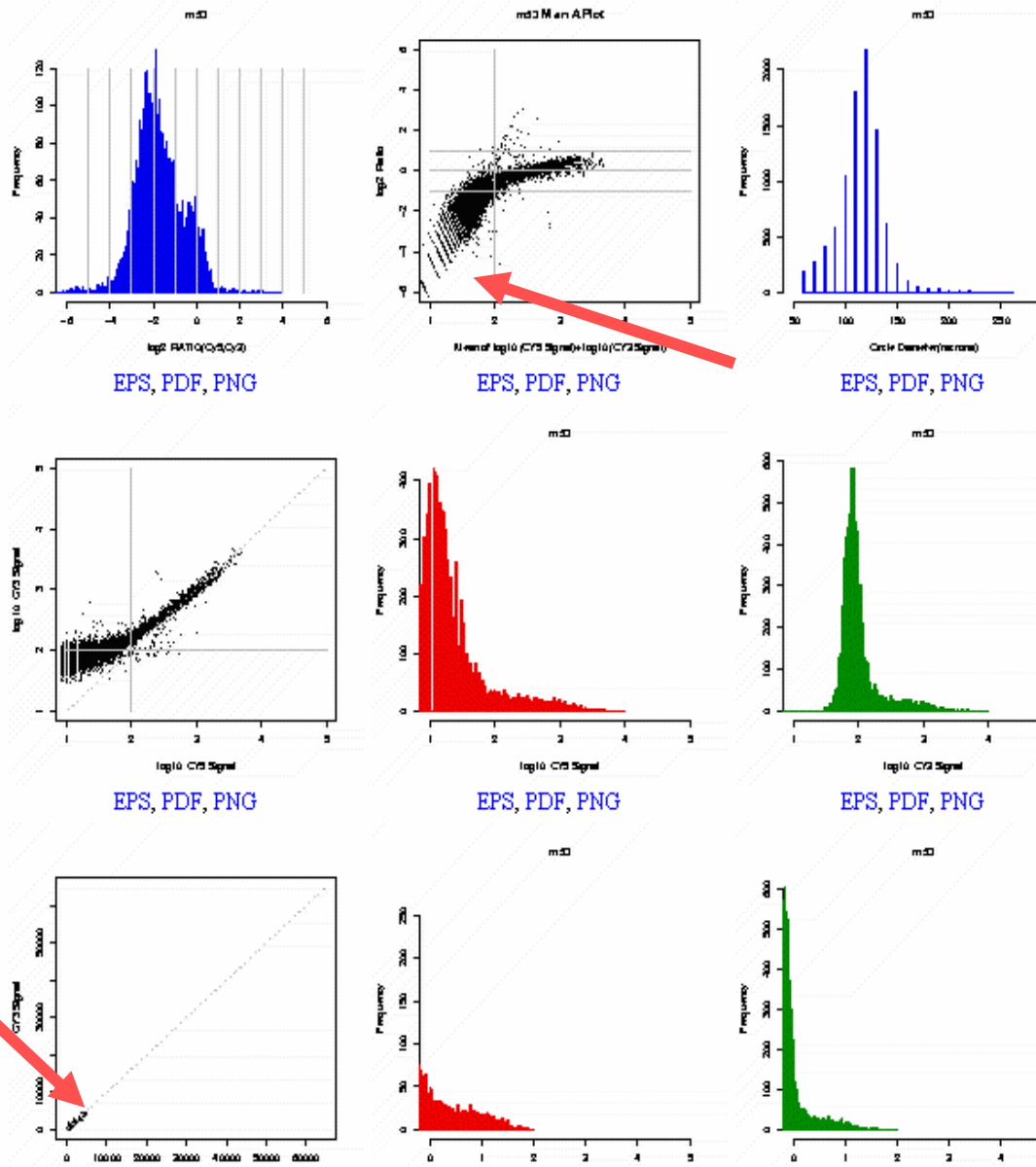


EPS, PDF, PNG

# Low Intensity Example

Signal Calculation: Mean Intensity - Median Background

Normalization Factor: 1.1821



- M vs A plot – ratio distribution dependent upon signal strength; see a “tail” toward green spots

- Spot sizes small
- Overall signal strength very weak – not a good range of signals on Cy3/Cy5 linear plot

- Bulk of red signals less than 10

- FYI, max signal is 65,000

# IV. Creating datasets and additional filtering

# **mAdb Analysis Paradigm:**

- 1. Create project; Upload arrays to that project**
- 2. Quality control – Project Summary and Graphical Reports**
- 3. Create a filtered dataset:**
  - Extract rows from database
  - Filter spots on quality parameters (spot size, S/N, etc.)
  - Normalize, so different arrays can be compared
  - Align genes from different array layouts (based on well IDs)
- 4. Apply Data/Gene criteria filters, if desired, to create subset dataset(s)**
- 5. Apply appropriate Analysis/Visualization Tools to the dataset(s)**
- 6. Repeat Steps 3, 4, and 5 as desired**
- 7. Interpret Datasets/Results**



# Lab 1 – Creating a filtered dataset

Goal: To use the Extended Extraction Tool to choose arrays from a project and filter on quality parameters.

Do NOT maximize the browser window, so multiple windows can be distinguished on the monitor.

# Lab 1. Choosing Project and Extended Dataset Extraction Tool

The screenshot shows the mAdb Gateway web page. At the top, there are links: Home Page, mAdb Gateway, Upload Status, Forums, Reference Info, Program Downloads, and GeneCards. Below these is the title "mAdb Gateway". A "NEW" tag is next to the "Upload" link. The text explains that the "Upload" link lists identifiers like Clone, Gene Symbol, LocusLink ID, UniGene ID, and Well ID, which can be used as filters with the Feature Properties Filtering tool. It instructs the user to choose one or more projects, select a tool, and click "Continue" to access previously extracted data in the "ncidemo's: Permanent" area. There are two dropdown menus: "Projects:" and "Tool:". The "Projects:" dropdown is open, showing a list of projects: "AX guest - Time Course Demo Set #1", "AX guest - Time Course Demo Set #2", "AX guest - Repeats and Reciprocal Retests Demo Set #3", "AU guest - Multiple Types Demo Set #4" (which is highlighted), "AU ncidemo - my project", and "AU ncidemo - Oligo and cDNA". A note below the dropdown states: "Note: Tools marked with '\*' only support selection of one project". The "Tool:" dropdown is set to "Extended Dataset Extraction". At the bottom is a "Continue" button. Three numbered circles (3, 4, 5) are overlaid on the image to indicate the steps: circle 3 points to the "Projects:" dropdown, circle 4 points to the "Tool:" dropdown, and circle 5 points to the "Continue" button.

Home Page | mAdb Gateway | Upload Status  
Forums | Reference Info | Program Downloads | GeneCards

## *mAdb Gateway*

**NEW** [Upload](#) lists of identifiers such as Clone, Gene Symbol, LocusLink ID, UniGene ID and Well ID. These lists can be used as filters with the Feature Properties Filtering tool.

Choose one or more Projects, select a Tool and Continue or access previously extracted data located in **ncidemo's:** [Permanent](#) area

**Projects:**

- AX guest - Time Course Demo Set #1
- AX guest - Time Course Demo Set #2
- AX guest - Repeats and Reciprocal Retests Demo Set #3
- AU guest - Multiple Types Demo Set #4**
- AU ncidemo - my project
- AU ncidemo - Oligo and cDNA

Note: Tools marked with "\*" only support selection of one project

**Tool:** Extended Dataset Extraction

Continue

1. Open a web browser and type the URL for the mAdb home page, <http://madb-training.cit.nih.gov>.

2. Click the first bullet on the home page, to access the **mAdb Gateway**, web page, shown at left. You will need to login the mAdb Gateway with the mAdb account as instructed.

3. On the mAdb Gateway Web page, in the **Projects:** list, select the "**guest – Multiple Types Demo Set #4**" project  
NOTE: You can select multiple projects by holding down the **Ctrl** key when you click on a project

4. On the **Tools:** menu just below, select "**Extended Dataset Extraction**"

5. Press the **Continue** button

# Lab 1. Selecting Filtering Options

**GenePix Extraction**

Note the --? marks items which lead to additional help when clicked

**Signal, Normalization & Ratio Options** --?

Signal Calculation: Median Int - Median Bkg

Normalization Method: 50th Percentile (Median)

Default Ratio: ChanB/ChanA (Cy5/Cy3)

☐ Limit Normalization to HouseKeeping Genes

Caution: Most array prints do not have an identified set of HouseKeeping Genes

☐ Include Control Features in the extracted set

1

1. In the **Signal, Normalization, & Ratio Options** panel, choose **Signal Calculation: Median Int – Median Bkg**, **Normalization Method: 50<sup>th</sup> Percentile (Median)**, and **Default Ratio: ChanB/ChanA**. Leave the checkboxes empty. Using this Normalization method, the output is re-normalized based on the spots which pass the filters.

**Spot Filter Options** --?

Check boxes on the left to activate specific criteria

☒ Exclude any Spots Indicated as Bad or Not Found

☒ Target diameter is between 50  $\mu$ m and 300  $\mu$ m

	Chan A (cy3)	Chan B (cy5)
<input type="checkbox"/> Target Pixels 1 SD above Bkg >=	50 %	50 %
<input type="checkbox"/> Signal Above Background >=	0 SDs	0 SDs
<input checked="" type="checkbox"/> Signal/Background Ratio >=	2	2
<input type="checkbox"/> Signal >=	100	100
<input checked="" type="checkbox"/> Override if Chan B Signal >=		5000
<input checked="" type="checkbox"/> Override if Chan A Signal >=	5000	
<input type="checkbox"/> Set Signal Floor Value =	100	100

2


2

2. In the **Spot Filter Options** panel, check the boxes on the left to activate the appropriate filter(s), and choose appropriate values by typing in numbers into the form elements to the right of each filter checkbox. For the purposes of this exercise, check:
  - Exclude any Spots indicated as **Bad or Not Found**
  - Target diameter is between **50**  $\mu$ m and **300**  $\mu$ m
  - Signal/background Ratio >= **2** and **2**
  - Override if Chan B Signal >= **5000**
  - Override if Chan A Signal >= **5000**

3. Go to next page of lab to choose arrays




## Lab 1. Selecting Dataset Properties and Arrays

**Dataset Properties** 

Rows Ordered by:

Dataset Location:

Dataset Label:

**Array Selection** 

[-] [A] [Submit] **3**

	A	1/R	mAdbID: Array Name & Short Description
<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28738: Mm-Incyte-v1p1-6 Sample 1/Type B
<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28739: Mm-Incyte-v1p1-7 Sample 2/Type B
<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28740: Mm-Incyte-v1p1-8 Sample 3/Type B
<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28741: Mm-Incyte-v1p1-9 Sample 4/Type B
<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28742: Mm-Incyte-v1p1-10 Sample 5/Type B
<input type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28733: Mm-Incyte-v1p1-1 Sample 1/Type A
<input type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28734: Mm-Incyte-v1p1-2 Sample 2/Type A
<input type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28735: Mm-Incyte-v1p1-3 Sample 3/Type A
<input type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28736: Mm-Incyte-v1p1-4 Sample 4/Type A
<input type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28737: Mm-Incyte-v1p1-5 Sample 5/Type A

1. In the **Dataset Properties** panel, choose **Rows Ordered by: Average(Log2 Ratio) and Descending; Dataset Location: Transient Area, and Dataset Label: “My Type A data – qual filtered”**.
2. In the **Array Selection** panel, choose just the Type A arrays using the radio buttons under **A**. **N.B.** If a dye swap or reverse fluor, check the **1/R** box to take the reciprocal value of the ratio for direct comparison.
3. Press **Submit**



# Lab 1. Waiting for Data Extraction ...

This page monitors the progress and allows you to continue when the results are available.

**Please wait for completion.**

Waiting ...

**Done! Please click**

**Continue**

**NOTE:** The dataset has been stored in your **Temporary** area. Datasets stored in the Temporary area are automatically deleted when 14 days expire with no access to the data. Accessing (that is "opening") the original set or a derived filtered/adjusted subset resets the "14 day clock". The mAdb Dataset management tool allows you to delete datasets from this area.

[Home](#) | [Analysis Tools](#) | [Forums](#) | [Reference Info](#) | [Program Downloads](#) | [GeneCards](#)

Intermediate screen which monitors the data extraction process. When the creation of the working dataset is complete, the user can continue to the Data Display page.

# Extended Tool: Signal, Normalization & Ratio Options:

- **Signal Calculation**

Mean Intensity – Median Background

Median Intensity – Median Background

- **Normalization**

- None

- 50<sup>th</sup> Percentile (Median)

Applied to extracted spots (spot filtered)

All spots or only Housekeeping spots (on limited prints)

- Pre-calculated 50<sup>th</sup> percentile (uses all data)

- Lowess non-linear normalization – in beta testing

- **Default Ratio**

Chan B/Chan A (CY5/CY3),

but for reverse fluor can choose Chan A/Chan B (CY3/CY5)

## Spot Filter Options:

- **Important** - Check box to Activate!
- Exclude any Spots Flagged as *Bad Or Not Found, Bad*
- Target diameter is between *default of 50 and 300 microns*
- **Target Pixels 1 Standard Deviation above background  $\geq N$  %**
- **Signal above background  $\geq N$  SDs (standard deviations)**
- **Signal/Background Ratio  $\geq N$**
- **Signal  $\geq N$  (raw signal intensities)**
- Override bracketed criteria ( in yellow above) if Chan B and /or A Signal  $\geq N$  (default is 5000)

# Signal Floor

- When one channel has a very low signal and the other has a moderate or high signal, the resulting ratio value could be misleading (i.e. very high/low)
- To adjust such a highly skewed ratio, mAdb allows the user to set a floor (e.g. 100) for signals below a threshold



# Lab 1. Main mAdb Dataset Display – Part 1

1. The listing at the top shows the array group, a link to the array image, a link to a histogram display, the re-calculated normalization factor (based on those spots which passed the quality filters), the array name, and the short description for all of the chosen arrays to be filtered
2. After the Dataset name (which can be **edited** with the link to the left), is the history of what was done in the preceding filtering step.
3. Go to the next page of the lab and scroll down to the bottom of the Web page.

**mAdb Dataset Display**

[View](#) Array Summaries

A 0.644 1. Mm-Incyte-v1p1-1 Sample 1/Type A

A 1.056 2. Mm-Incyte-v1p1-2 Sample 2/Type A

A 0.627 3. Mm-Incyte-v1p1-3 Sample 3/Type A

A 0.551 4. Mm-Incyte-v1p1-4 Sample 4/Type A





















A 0.727 5. Mm-Incyte-v1p1-5 Sample 5/Type A

[Edit](#) Data for Dataset: My Type A data - qual filtered

5 Arrays and 5276 Expression Rows extracted.  
Default Ratio: ChanB/ChanA (Cy5/Cy3)  
Signal calculation: Median Intensity minus Median Background  
Any Features designated Control were excluded.  
Normalization method: 50th Percentile (Median) using all spot filtered Genes  
Spot Filter Options:  
Include Spots not flagged BAD or Not Found  
AND Target diameter >= 50 um AND Target diameter <= 300 um  
AND Both Chan A and Chan B Signal/Background Ratios >= 2.000  
Override other Chan A & B criteria and Include if Chan A Signal >= 5000 OR Chan  
Data was extracted and aligned by the Inventory Well ID  
Any multiple occurrences of Well ID were reduced to a single instance  
by selecting the one with the strongest signal (Chan A + Chan B)

# Lab 1. Main mAdb Dataset Display – Part 2

Records 1 to 25 of 5276 total records

#1	#2	#3	#4	#5	Well ID	Feature ID	Description
					616842	<a href="#">IMAGE:481151</a>	procollagen, type IX, alpha 1
					617147	<a href="#">IMAGE:493658</a>	lipocalin 2
					614212	<a href="#">IMAGE:402800</a>	Mus musculus transcribed sequences
					614066	<a href="#">IMAGE:374725</a>	RIKEN cDNA 2310047E01 gene
					613588	<a href="#">IMAGE:333418</a>	protein tyrosine phosphatase, receptor type, 1
					617076	<a href="#">IMAGE:571759</a>	RIKEN cDNA 9530006B08 gene
					614354	<a href="#">IMAGE:403453</a>	protein tyrosine phosphatase, receptor type, 1
					619013	<a href="#">IMAGE:832158</a>	extracellular proteinase inhibitor




1. This is the main page to display expression data, and as we will see on the next page, is highly customizable. Each column represents an array, each row a gene feature. Grey boxes are either missing values or data that was filtered out due to low quality. You can page through the data using the **arrow** just above the columns of data.
2. The **Well ID** uniquely identifies the piece of DNA used on that feature, and the **Feature ID** is an accession number. The **Well ID** is a hyperlink to a montage of the spot images, whereas the **Feature ID** is a Hyperlink to a **Feature Report**, integrating information about the gene related to the feature and its function(s).
3. There is a brief description of the feature on the right hand side of the display. Note that each column can be sorted in either ascending or descending order using the **grey arrows** above each column.

Feature Report - Microsoft Internet Explorer

File Edit View Favorites Tools Help Links »

## *mAdb* Feature Report


---

<b>Clone</b>	<a href="#">IMAGE:301551</a>																																
<b>Library Source</b>	Soares_fetal_lung_NbHL19W																																
<b>Sequence Verification</b>	Unknown																																
<b>Annotated Simple PID</b>	Integrin, alpha V (vitronectin receptor, alpha polypeptide, antigen CD51)																																
<b>Annotated NG Assignment</b>	<a href="#">M14648</a> Human cell adhesion protein (vitronectin) receptor alpha subunit mRNA, complete cds																																
<b>Annotated Categories</b>	Adhesion																																
<b>5' Sequence</b>	<a href="#">W17002</a> UCSC's <a href="#">GenomeViewer</a>																																
<b>5' UG Title</b>	integrin, alpha V (vitronectin receptor, alpha polypeptide, antigen CD51)																																
<b>5' UG Cluster</b>	<a href="#">TP Hs.295726</a> NCBI's <a href="#">LocusLink</a> Stanford's <a href="#">S.O.U.R.C.E.</a>																																
<b>5' UG Gene</b>	ITGAV <a href="#">GeneCards</a> <a href="#">MedMiner</a> NCBI's <a href="#">Map Viewer</a>																																
<b>5' UG LL Summary</b>	ITAGV encodes integrin alpha chain V. Integrins are heterodimeric integral membrane proteins composed of an alpha chain and a beta chain. The I-domain containing integrin alpha V undergoes post-translational cleavage to yield disulfide-linked heavy and light chains, that combine with multiple integrin beta chains to form different integrins. Among the known associating beta chains (beta chains 1,3,5,6, and 8; 'ITGB1', 'ITGB3', 'ITGB5', 'ITGB6', and 'ITGB8'), each can interact with extracellular matrix ligands; the alpha V beta 3 integrin, perhaps the most studied of these, is referred to as the Vitronectin receptor (VNR). In addition to adhesion, many integrins are known to facilitate signal transduction.																																
<b>5' UG Ontology</b>	<table border="0"> <tr> <td><a href="#">GO</a>™ Annotations</td> <td><a href="#">Evidence</a> </td> <td>Source</td> <td>Pub</td> </tr> <tr> <td>• <a href="#">cell adhesion</a></td> <td>P</td> <td>Proteome</td> <td><a href="#">PM</a></td> </tr> <tr> <td>• <a href="#">cell adhesion receptor</a></td> <td>P</td> <td>Proteome</td> <td><a href="#">PM</a></td> </tr> <tr> <td>• <a href="#">integral plasma membrane protein</a></td> <td>P</td> <td>Proteome</td> <td><a href="#">PM</a></td> </tr> <tr> <td colspan="4">Other Annotations</td> </tr> <tr> <td>• Integral membrane</td> <td>NR</td> <td>Proteome</td> <td><a href="#">PM</a></td> </tr> <tr> <td>• Receptor (signalling)</td> <td>NR</td> <td>Proteome</td> <td><a href="#">PM</a></td> </tr> <tr> <td>• Control of Cell Proliferation</td> <td>E</td> <td>Proteome</td> <td><a href="#">PM</a></td> </tr> </table>	<a href="#">GO</a> ™ Annotations	<a href="#">Evidence</a> 	Source	Pub	• <a href="#">cell adhesion</a>	P	Proteome	<a href="#">PM</a>	• <a href="#">cell adhesion receptor</a>	P	Proteome	<a href="#">PM</a>	• <a href="#">integral plasma membrane protein</a>	P	Proteome	<a href="#">PM</a>	Other Annotations				• Integral membrane	NR	Proteome	<a href="#">PM</a>	• Receptor (signalling)	NR	Proteome	<a href="#">PM</a>	• Control of Cell Proliferation	E	Proteome	<a href="#">PM</a>
<a href="#">GO</a> ™ Annotations	<a href="#">Evidence</a> 	Source	Pub																														
• <a href="#">cell adhesion</a>	P	Proteome	<a href="#">PM</a>																														
• <a href="#">cell adhesion receptor</a>	P	Proteome	<a href="#">PM</a>																														
• <a href="#">integral plasma membrane protein</a>	P	Proteome	<a href="#">PM</a>																														
Other Annotations																																	
• Integral membrane	NR	Proteome	<a href="#">PM</a>																														
• Receptor (signalling)	NR	Proteome	<a href="#">PM</a>																														
• Control of Cell Proliferation	E	Proteome	<a href="#">PM</a>																														
<b>5' UG RefSeq</b>	<a href="#">NM_002210</a>																																
<b>5' UG Cytoband</b>	2q31-q32																																
<b>5' Submitted PID</b>	gb:M14648 VITRONECTIN RECEPTOR ALPHA SUBUNIT PRECURSOR (HUMAN);																																

Internet



# Lab 1. Main mAdb Dataset Display – Part 3

**Dataset Retrieval & Display Options** 

Dataset formatted for Eisen Cluster **2**

---

☒ Show Array Details at the top of the page

Background Color Red/Yellow/Green Contrast 2 **1**

Limiting display to to 25 genes

<input checked="" type="checkbox"/> Show Data Values	<input type="checkbox"/> Use Names in Column Heading
<input checked="" type="checkbox"/> Apply log2 transform	<input type="checkbox"/> Use Description in Column Heading
<input checked="" type="checkbox"/> Show Spot Images	<input type="checkbox"/> Show Gene Symbols
<input type="checkbox"/> Show Map Information	<input type="checkbox"/> Show UniGene Cluster
<input type="checkbox"/> Show BioCarta Pathways	<input type="checkbox"/> Show KEGG Pathways
<input type="checkbox"/> Show GO Tier 2 Component	<input type="checkbox"/> Show GO Tier 3 Component
<input type="checkbox"/> Show GO Tier 2 Function	<input type="checkbox"/> Show GO Tier 3 Function
<input type="checkbox"/> Show GO Tier 2 Process	<input type="checkbox"/> Show GO Tier 3 Process
<input checked="" type="checkbox"/> Show Gene Description	<input type="checkbox"/> Show GO Terms
<input type="checkbox"/> Show Average(Log2 Ratio)	<input type="checkbox"/> Show Max(Log2 Ratio)-Min(Log2 Ratio)
<input type="checkbox"/> Show Variance	

[Save](#) a Feature Property List (used with the Feature Properties Filtering tool).

1. Here is where the data display on the preceding page can be customized, by checking or unchecking the checkboxes next to each column name. One can include numerical summary data (**Average(Log2 Ratio)**, **Variance**, **Max(Log2 Ratio)-Min(log2 Ratio)**); pathways (**KEGG**, **BioCarta**); Genome Ontology (**GO**) classifications; and display individual **Spot Images**, among others. One can also change or eliminate the **Background Color** on the table of data values, adjust its **Contrast** (the point where max red and green are reached), and also adjust how many genes are displayed in the table on a Web page (the default is 25). Once the choices are made, push the **Redisplay** button to refresh the page with your desired changes.
2. You can also retrieve the dataset for MS-Excel, the Eisen Cluster program format, or in tab-delimited files for the Macintosh, PC, or UNIX platforms.



# Lab 1. Main mAdb Dataset Display – Part 4



1. Once the data is filtered by quality, the most likely next step is to do additional filtering and create a subset of this parent dataset. Under *Filtering/Grouping/Analysis Tools*, choose the default pulldown option of **Additional Filtering Options** and press **Proceed**.
2. Alternately, one could access *Interactive Graphical Viewers* from here, but we will return and demonstrate that later.
3. Also, you could **Access other Datasets in your Transient Area** from here with the link above the yellow panels.

# Affy Extraction Tool (for Absolute data)

## Affymetrix Absolute Extraction

Note the ⓘ marks items which lead to additional help when clicked

### Data Transformation Options ⓘ

Transformation: **Centered to scale target 500** ▼

☐ Signal Floor =

### Filter Options ⓘ

Check boxes on the left to activate specific criteria  
▼

- ☐ Exclude All Present (P) Calls
- ☐ Exclude All Marginal (M) Calls
- ☐ Exclude All Absent (A) Calls

- 
- ☐ Present (P) Call AND Signal  $\geq$
  - ☐ Marginal (M) Call AND Signal  $\geq$
  - ☐ Absent (A) Call AND Signal  $\geq$

# Break


# V. Basic data analyses and dataset management



# Tools for Basic Data Analyses

- Once you have a filtered dataset, you can:
  - Filter further for missing values and/or gene ratio levels
  - Do an *ad hoc* Keyword search
  - Filter datasets by lists of gene identifiers
  - View GO and Pathway Summaries
  - Use Graphic Tools
    - Interactive Scatter Plot
    - Correlation Summary Report
    - Multiple Array Viewer

## Additional Filtering and Analysis Options

Additional Data Filtering/Adjustment/Analysis 

Choose an analysis option Additional Filtering Options and Proceed

Choose a Graphical View ▼ and View

Retrieve Data Set format ▼

Retrieve Data set format ☒ Apply log2

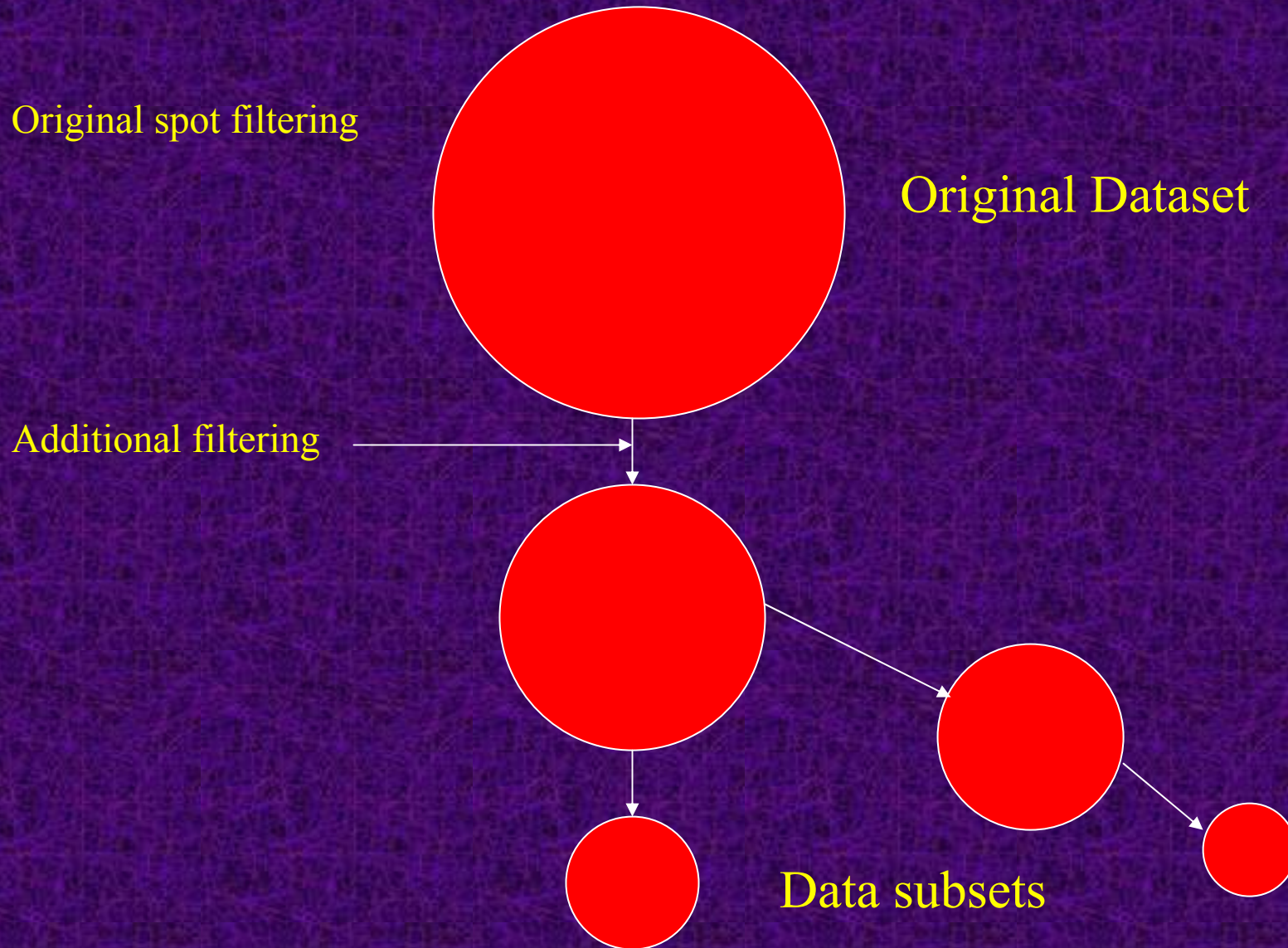
Redisplay ☒ Show Array Details at the top of the page

Background Color Red/Yellow/Green Contrast 1

Limiting display to to 25 genes

- Additional Filtering Options
- Ad Hoc Query/Filtering Options
- Feature Property Filtering Options
- Array Order Designation/Filtering
- Array Group Assignment/Filtering
- Filter/Group by Array Properties
- Average Arrays within Groups
- Two or more Group Comparison
- PAM: Prediction Analysis for Microarrays
- Boolean Comparison with another Set
- Clustering: Hierarchical
- Clustering: Kmeans
- Clustering: SOM
- Correlation Summary Report
- Gene Ontology Summary Report
- Pathways Summary Report
- Save As a New Dataset

# Dataset Structure -Filtering hierarchy /tree structure



# Dataset History

12 Arrays and 3877 Expression Rows extracted with Spot Filter Options:  
Chan A Spot size (percentage)  $\geq 25$  and Chan B Spot size (percentage)  $\geq 25$   
AND ( ( Chan A Signal  $\geq 50$  AND Chan B Signal  $\geq 50$  ) OR ( Chan B Signal  $\geq 2500$  ) )  
AND ( Spots not flagged BAD or Not Found )  
Note: For all GenePix results from Axon scanned arrays Chan A is CY3 and Chan B is CY5.  
Rows ordered by Maximum(Ratio)/Minimum(Ratio) descending.

---

Thu Oct 4 09:22:48 EDT 2001

Interactive Array Filtering  
12 arrays and 3877 genes in the [original dataset](#)  
6 arrays and 3875 genes in the output data set.  
6 Arrays were interactively excluded  
2 Genes excluded for having zero observations in the resultant array set.

---

Thu Oct 4 09:23:56 EDT 2001

6 arrays, 3875 genes in the [input dataset](#)  
941 Genes and 6 arrays passed filters  
111 genes excluded for being present in less than 80% arrays.  
2823 genes excluded where variance is in the lowest 75 percentile (Variance $<0.44$ ).

Link to the [output Dataset](#)

A log is maintained for each dataset tracing the analysis history.  
When the history is displayed, links are provided to allow the user to recall any dataset in the analysis chain.



# Lab 2 –Additional Filtering

Goal Lab 2: To filter out missing values for genes and to look for genes up or down regulated at least 2-fold.

- Applies selected filtering options to the dataset based on values in the data and creates a new subset.
- For gene filters, ratios are expressed as fold changes and all calculations are done in log space

# Lab 2. Additional Filtering

**Data Filtering Options**

Check boxes on the left to activate specific filters

**Missing Value Filters**

☒ Genes: Require values in  $\geq$  90 % of Arrays

☐ Arrays: Require values in  $\geq$  70 % of Genes

**Gene Filters**

☒ Ratio  $\geq$  2 in  $\geq$  80 % of Arrays  
☒ Apply Symmetrically

☐ Ratio  $\geq$  3 in  $\geq$  5 Arrays OR  
 Ratio  $\leq$  0.2 in  $\geq$  5 Arrays

☐ Average Ratio  $\geq$  2  
☒ Apply Symmetrically

☐ Max (Ratio) / Min (Ratio)  $\geq$  3

☐ Variance (Gene Vector) percentile  $\geq$  70 %

Subset Label: My Type A data - mv & 2x

**Filter** **Cancel**

Home Page | mAdb Gateway | Upload Status  
 Forums | Reference Info | Program Downloads | GeneCards

NIH Bioinformatics support provided by  
 BIMAS/CIT.  
 We can be contacted by email.

1. Filter the rows of data from the parent dataset for missing values, requiring genes in  $\geq$  **80%** of Arrays. Alternately, it is possible to filter out Arrays by requiring values in  $\geq$  70% of genes, for example.
2. mAdb supports a wide variety of Gene Filters: We will use a **Ratio  $\geq$  2 in  $\geq$  80 % of Arrays**, with the **Apply Symmetrically** box checked to obtain genes up and down-regulated by 2-fold or more.  
 Other options are:
  - Filter for at least 3 fold up in 5 or more arrays OR at least 5-fold down (0.2x up) in 5 or more arrays.
  - Filter for an average Ratio across the row at least two fold or more, applied symmetrically to obtain genes with an average ratio two-fold or more up or down regulated.
  - Filter for those rows showing a difference between the maximum ratio and minimum ratio on each row of 3 fold or more
  - Rank the genes by percentile of variance, and then filter for those genes in the top 30%ile of variance – ie. The genes that vary the most across the rows statistically.
  - N.B. Filters are applied in order from top to bottom on this page
3. Label the subset **“80% missing values & 2 fold up/down”** to reflect what you did .
4. Press the **Filter** button to continue and create the desired subset.

# Lab 2. Additional Filtering

## mAdb Dataset Display

[View](#) Array Summaries

[Edit](#) Data for Subset: **filter**  
from Dataset: **test for class**

The filter input data set contained 5 arrays and 5276 genes.  
The filtered output data set contains 5 arrays and 340 genes.  
3122 genes excluded for being present in less than 80% (4) arrays.  
1814 genes excluded by `abs(log2) >= 2` or `<= 0.50` in at least 80% (4) array(s).

View the complete [History](#).

[Expand](#) this Dataset.

Access Datasets in your [Temporary](#) area.

Records 1 to 25 of 181 total records displayed.

A	A	A	A	A					
#1	#2	#3	#4	#5	Aver	Well ID	Feature ID	Description	
1.0638	0.3546	1.2041	1.0461	1.3991	1.0135	613095	IMAGE:894372	RIKEN cDNA 2400003B06 gene	
-1.4357	-1.9626	-1.5519	-1.4536	-1.6412	-1.6090	613104	IMAGE:805218	UDP-glucose pyrophosphorylase	
-2.2922	-2.3579	-2.1242	-1.8827	-1.8532	-2.1020	613113	IMAGE:775189	isocitrate dehydrogenase 3 (NAD	
0.9825	1.0011	1.1738	1.2341	1.2136	1.1210	613221	IMAGE:792954	protein phosphatase 2, regulatory	
1.7611	0.7633	1.4790	1.4585	1.6127	1.4149	613358	IMAGE:961282	CDC28 protein kinase 1	
1.3170	0.2659	1.1725	1.5469	1.6639	1.1932	613381	IMAGE:317466	DNA segment, Chr 7, Wayne Str	
-2.7152	-2.8092	-2.2288	-2.2725	-2.6602	-2.5372	613412	IMAGE:331768	Unknown	
1.3435	1.3554	1.2086	1.1099	0.9873	1.2010	613459	IMAGE:335112	destrin	

1. Note that in the returned dataset, there are many fewer missing values – see the history log for how many genes were filtered out to create this subset.
2. This is a data subset – you can view the complete History of the dataset via this link.
3. You can also **Expand this Dataset** to show the parent and all children, or again **Access Datasets in your Transient Area** via these links.

### Notes:

- Applies selected filtering options to the dataset based on values in the data and creates a new subset.
- For gene filters, ratios are expressed as fold changes and all calculations are done in log space

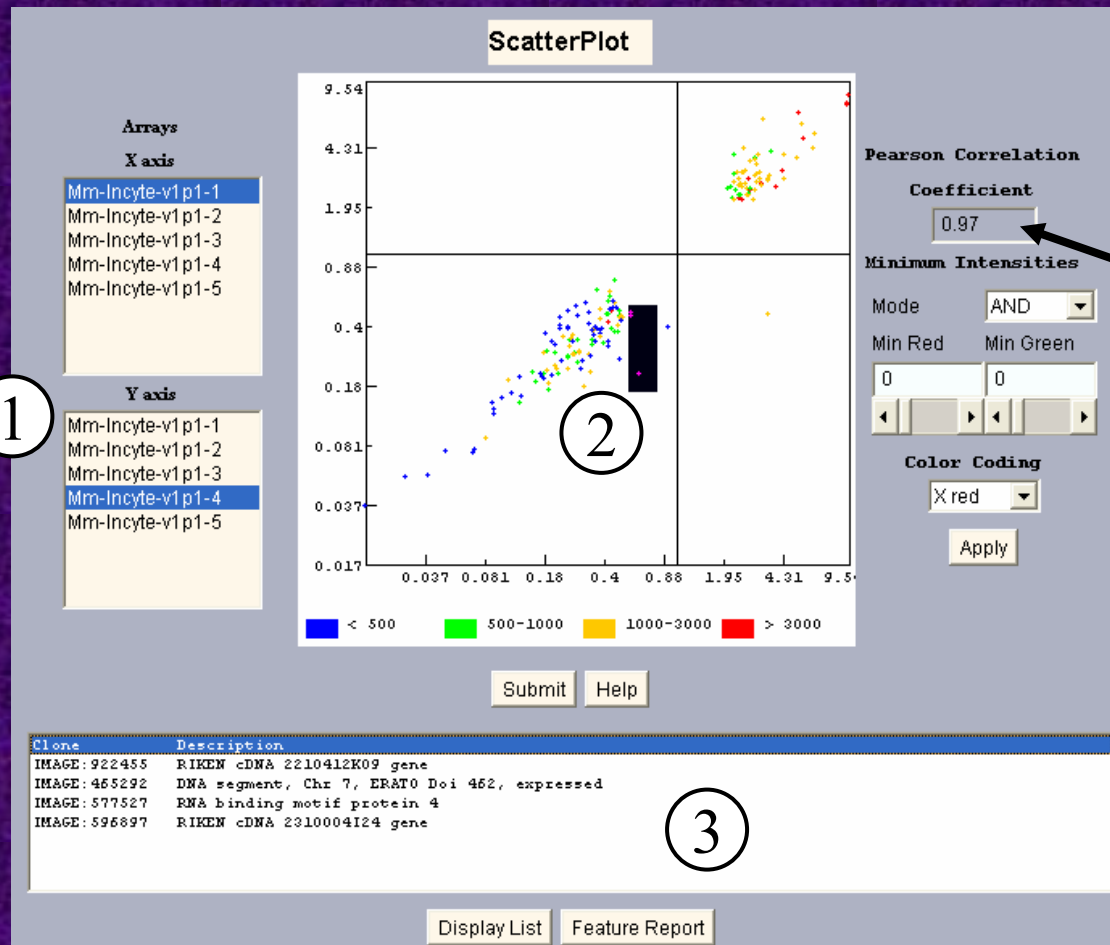


# Lab 3 – Using More Tools

Goals Lab 3: To use the Interactive Scatter Plot, Correlation Summary Report, Ad Hoc query, Multiple Array Viewer, Pathway, GO summary Reports and feature properties filtering,



# Visualization Tools – Interactive Scatter Plot Applet



•Replicate experiments should be on a 45° angle (slope of 1) and the Pearson Correlation Coefficient should be approaching 1

•Reverse fluor experiments should have a Pearson Correlation Coefficient approaching -1

Access from *Interactive Graphical Viewers* Menu on main **mAdb Dataset Display** page:

1. Choose Arrays to be compared on X and Y axes
2. Can select outlying spots with mouse – genes will be shown in window below plot
3. Can get **Feature Report** by clicking on gene name in lower display box

# Correlation Summary Report

Allows pair wise comparison  
of all arrays in a project –  
useful for comparing  
replicates and reverse fluors

## mAddb Correlation Report

[View](#) Array Summaries

[Return](#) to the input dataset.











[Redisplay](#) Background Color Scheme

Color Saturation Max/Mid/Min

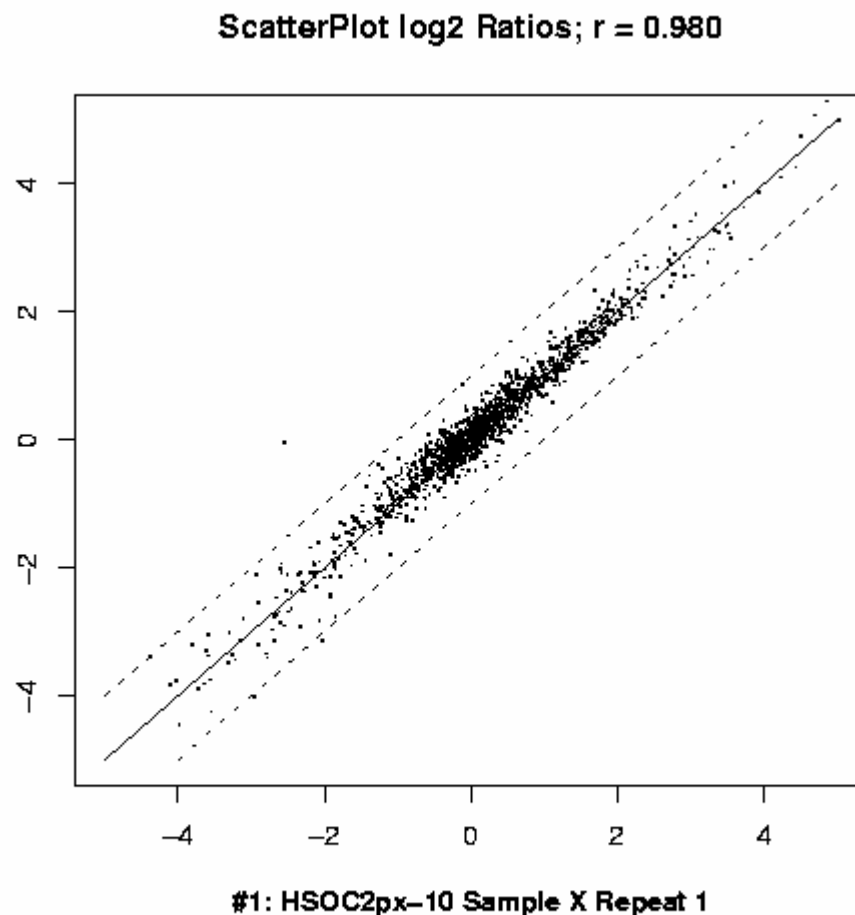
Note: For proper coloring Max > Mid > Min

Note: Click on the Correlation values to display the corresponding ScatterPlot

### Correlations


	A	A	A	A*	A*					
	#1	#2	#3	#4	#5	Grp		Array Name	Array	
1.	-	0.948	0.980	0.727	0.711	A	 	1. HSOC2px-10	Samp	
2.		-	0.928	0.742	0.738	A	 	2. HSOC2px-11	Samp	
3.			-	0.764	0.756	A	 	3. HSOC2px-12	Samp	
4.				-	0.972	A*	 	4. HSOC2px-14	Samp	
5.					-	A*	 	5. HSOC2px-15	Samp	

#3: HSOC2px-12 Sample X Repeat 3



[Click to close](#)

# Selecting Ad Hoc Query Tool

Filtering/Grouping/Analysis Tools 

Choose a Tool  and

Choose a View  and

Dataset for

☒ Show Background Limiting

- Ad Hoc Query/Filtering Options
- Additional Filtering Options
- Ad Hoc Query/Filtering Options
- Feature Property Filtering Options
- Array Order Designation/Filtering
- Array Group Assignment/Filtering
- Filter/Group by Array Properties
- Average Arrays within Groups
- Two or more Group Comparison
- PAM: Prediction Analysis for Microarrays
- Boolean Comparison with another Set
- Clustering: Hierarchical
- Clustering: Kmeans
- Clustering: SOM
- Correlation Summary Report
- Gene Ontology Summary Report
- Pathways Summary Report
- Save As a New Dataset

Select “Ac

## mAdb Ad Hoc Query

Check boxes on the left to activate additional Ad Hoc filters

▼

① Gene Description Contains receptor

② ☒ and Chromosome Begins with 4

③

Subset Label: My Type A Ad Hoc Query - receptor & chr 4

Filter Cancel

Basically a Boolean Keyword search; access from main **mAdb Dataset Display** page tool pulldown menu:

1. Can pick from **BioCarta Pathway, Feature ID, Gene Description, Gene Symbol, GO term, KEGG Pathway, Map Location, UniGene ID, Well ID**
2. Check box to add another term with **AND/OR** choice
3. Can choose **Contains, Begins With, Equals, Does Not Contain, Does Not Begin With, Does Not Equal** for search qualifier



# Output of Ad Hoc Query

## mAdb Dataset Display

[View](#) Array Summaries

[Edit](#) Data for Subset: **My Type A Ad Hoc Query - receptor & chr 4**  
from Dataset: **test for class**

### Ad Hoc Filtering

5 arrays and 340 genes in the input dataset











5 arrays and 2 genes in the output dataset.

### Ad Hoc Filter:

Gene Description Contains 'receptor'

AND Chromosome Begins with '4'

Records 1 to 2 of 2 total records displayed.

A	A	A	A	A						
#1	#2	#3	#4	#5	Aver	Well ID	Feature ID	Map	Description	
 3.2349	 2.7053	 3.0574	 2.8567	 3.3126	3.0334	614354	IMAGE:403453	4 C6-D1	protein tyrosine phosphatase, receptor type, F	
 -1.9201	 -2.4286	 -1.7173	 -1.9279	 -1.8618	-1.9711	620446	IMAGE:735186	4 D2.3	nuclear receptor binding factor 1	

## Pathway Summary Report

Total number of features: 97

Total number of features mapped to a KEGG Pathway: 8

Total number of features mapped to a BioCarta Pathway: 5

Total number of features not mapped to any Pathway: 84

**NOTE:** Clicking on # of features creates a new subset containing only the features the mapped to the Pathway.

**NOTE:** Clicking on BioCarta Pathway ID displays the pathway.

# of Features	BioCarta Pathway
1	m_cxcr4Pathway
1	m_ifngPathway
1	m_keratinocytePathway
1	m_etsPathway
1	m_ccr5Pathway
1	m_nktPathway
1	m_malatePathway
1	m_th1th2Pathway
1	m_eel1Pathway

**NOTE:** Clicking on # of features creates a new subset containing only the features the mapped to the Pathway.

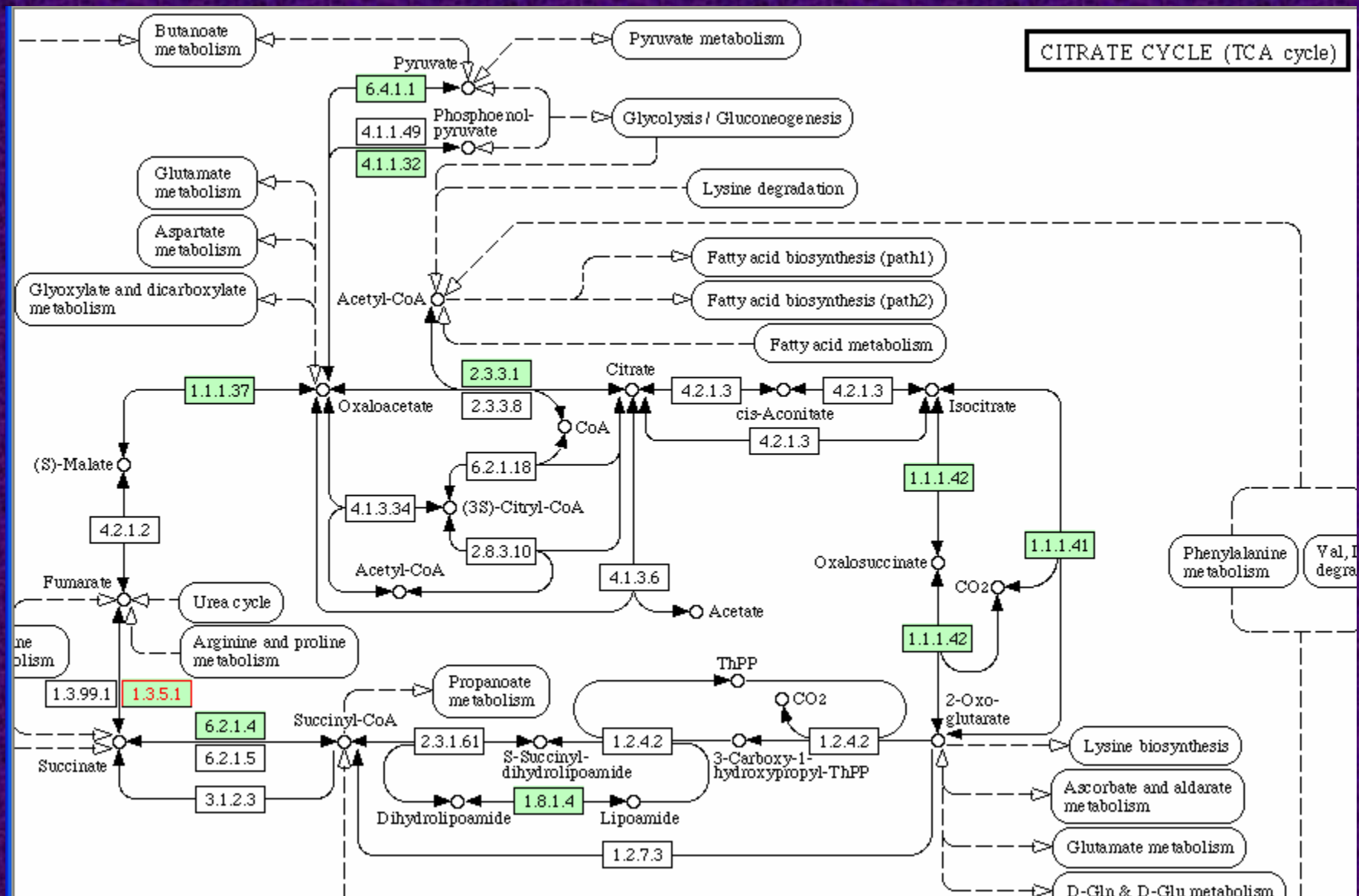
**NOTE:** Clicking on KEGG Pathway ID displays the pathway with features high lighted.

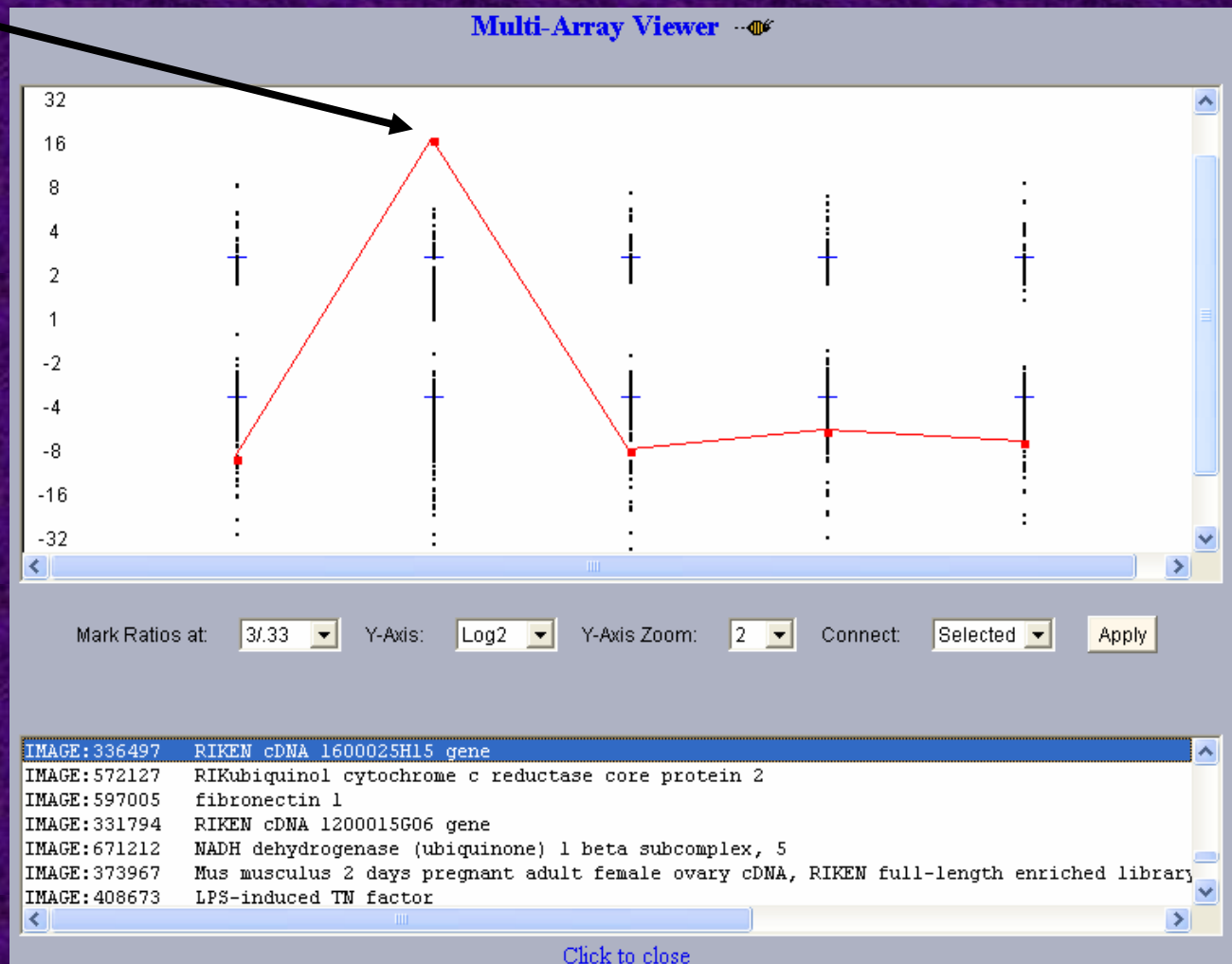
# of Features	KEGG Pathway
2	mmu00561
2	mmu00190
1	mmu00193
1	mmu00362
1	mmu00710
1	mmu00020

Access from main tools menu on **mAdb Dataset Display Page:**

1. Clicking on **# of Features** link creates a new dataset of just those features.
2. Clicking on **BioCarta Pathway** links show pathway on BioCarta Web site.
3. Also have **GO Ontology Summary Report**.

# A KEGG Pathway





Access from *Interactive Graphical Viewers* Menu on main **mAdb Dataset Display** page :

1. Can choose a point on graphical window to display a graph of that gene's expression which passes through that point
2. Can select a gene name on lower list and graph will appear in plot above
3. Can get **Feature Report** by clicking on gene name in lower display box



# Save Feature Property List from mAdb Data Display

☐ Show GO Tier 2 Process    ☐ Show GO Tier 3 Process  
☒ Show Gene Description    ☐ Show GO Terms  
☒ Show Average(Log2 Ratio)    ☐ Show Max(Log2 Ratio)-Min(Log2 Ratio)  
☒ Show Variance

**Save** a Feature Property List (used with the Feature Properties Filtering tool).

→ Records 1 to 25 of 181 total records displayed.


A	A	A	A	A							
#1	#2	#3	#4	#5	Aver	Var.	Well ID	Feature ID	Gene	Descripti	
3.2349	2.7053	3.0574	2.8567	3.3126	3.0334	0.052	614354	IMAGE:403453	Ptprf	protein tyr	
3.2446	2.5069	3.0894	2.8202	2.1709	2.7664	0.152	614891	IMAGE:421150	Cd24a	CD24a an	

- Can save a list of well IDs, clone/feature identifiers, gene symbols, UniGene identifiers from the dataset display page
- List can be stored as local to the dataset or globally available to all datasets

# Save Feature Property List from mAdb Data Display

**mAdb: Save a Feature Property List**

---

Feature Property List 

Save a List of: mAdb Well IDs ▼

Store the List as: Local (Available only in this Dataset) ▼

List Label:

☐ Overwrite any existing list with the same label

---

Save

- Can save a list of well IDs, clone/feature identifiers, gene symbols, UniGene identifiers from the dataset
- List can be stored as local to the dataset or globally available to all datasets

# Or Manually Create a List of Identifiers for Filtering

## mAdb Identifiers List Upload

This Form allows you to upload a list of Identifiers such as Clone, UniGene, Well ID. Uploaded lists are available as filter options in the "Feature Properties Filtering Tool".

Note; There is no need to specify the type of identifier in the "List Label". The system remembers each type of list presents your lists segregated and identified by type.

Type of List:	Clone/Feature Identifier (IMAGE:12345, 12345_at) ▼
List Label:	Rab clones
	<input type="checkbox"/> Overwrite an existing list with the same label
Paste/Type in List: (One element/line)	<div>IMAGE: 619501 IMAGE: 466099 IMAGE: 779604</div>
<div>Clone/Feature Identifier (IMAGE:12345, 12345_at) Gene Symbol (BRCA1) LocusLink Identifier (12345) UniGene Identifier (X.1234) mAdb Well ID (12345)</div>	
<div>Submit</div>	

Can paste in list of identifiers; must use format as shown in pull down menu

# Additional Filtering by Feature Properties and/or Lists

**Feature Properties Filtering Options**

Check boxes on the left to activate specific filters

☐ Include only Designated Housekeeping Genes

☐ Include only Designated Control Features

☐ Include only where Well ID = 0

☐ Include only where 0 <= Well ID <= 96

☒ Include only where Clone is in **rab clones**

☐ Include only where UniGene is in mouse unigene ID

☐ Include only where Well ID is in mylist

\*\*\* indicates lists local to this dataset

Subset Label: rab clone list filtering

**Filter** **Cancel**

Filters another dataset so that only those clones matching feature properties or in selected lists are returned



Records 1 to 3 of 3 total records displayed.

A	A	A	A	A	B	B	B	B	B					
#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Aver	Well ID	Feature ID	Description	
1.4976	1.1814	1.6701	1.4505	1.4512	0.3770	0.4494	1.0830	1.1677	1.1442	1.1472	615892	IMAGE:466099	RAB6, member RAS oncogene family	
	0.5956	0.1952	0.5601	0.2995	-0.2228	-0.0173	0.1918	0.0718	0.3550	0.2254	618176	IMAGE:619501	RAB1, member RAS oncogene family	
0.2911	0.4528	0.3646	0.3658	-0.1094	-0.2409	-0.0061	-0.1560	0.2725	-0.0365	0.1198	613367	IMAGE:779604	RAB7, member RAS oncogene family	



# Managing Feature Lists

## Manage Feature Identifier Lists

[Need Help?](#) 

Check boxes to select Identifier lists to Delete

▼ List (Click on a List to View/Edit) List type

<input type="checkbox"/>	<a href="#">Esther's list</a>	Clone
<input type="checkbox"/>	<a href="#">my favorite genes</a>	Clone
<input type="checkbox"/>	<a href="#">my interesting list</a>	Clone
<input type="checkbox"/>	<a href="#">list of 340 genes 2x up down</a>	Gene
<input type="checkbox"/>	<a href="#">receptors on chrom 5</a>	Gene
<input type="checkbox"/>	<a href="#">oxidative phosph</a>	UniGene
<input type="checkbox"/>	<a href="#">PAM-unigene</a>	UniGene
<input type="checkbox"/>	<a href="#">mylist</a>	Well ID

Delete



## Feature Identifiers List

Retrieve Esther's list formatted for PC

Type of List: Clone

Original List Label: Esther's list

List Label: Esther's list

List Values:  
(1 item per line)

IMAGE: 697383  
IMAGE: 790571  
IMAGE: 920235  
IMAGE: 466099  
IMAGE: 316187  
IMAGE: 333232  
IMAGE: 762516  
IMAGE: 400592  
IMAGE: 467790  
IMAGE: 463386

List Value Order is maintained

Save

Cancel

# Lab 4 – Dataset Management

Goal: To manage datasets by renaming, moving its storage location or deleting them.

## Accessing Temporary Datasets


1

**Manage** datasets located in your: [Temporary](#) or [Permanent](#) area

2

Switch to **accessing** datasets located in your: [Permanent](#) area

3

Temporary Datasets		Created		Containing		Need Help? 		Gene Information	
				Arrays	Genes			Refreshed	
<a href="#">Edit</a>	hands-on qual filter	Dec 12	11:37:02am	5	5276	<a href="#">Open</a>	<a href="#">Expand (1)</a>	<a href="#">Refresh</a>	Dec 12 11:38:27am

4

5

### Dataset Access:

1. Can **Manage** Transient, Temporary, or Permanent Areas (wait for next slide)
2. Can **Access** other dataset areas which contain data (i.e. Permanent)
3. Can **Edit** dataset name
4. Can **Expand** to see parent dataset and all children of that parent
5. Can **Refresh** Gene Information – see next slide

# Refreshing Gene Information

- Clicking “refresh” link updates all of the gene information in the dataset (UniGene cluster, Description, Pathway info, Map info...)
- May want to “Save as a New Dataset”, and then refresh, if you don’t want gene names to change as you near publication



# Save as New Dataset

**mAdb Dataset Display**

[View](#) Array Summaries

[Edit](#) Data for Subset: **class 1** - **900%**  
from Dataset: **class 1/27 - 900%**

The filter input data  
The filtered output data  
3122 genes excluded from  
1814 genes excluded from

View the complete [History](#).

[Expand](#) this Dataset.  
Access Datasets in your [Ter](#)

Choose a Tool

- Additional Filtering Options
- Ad Hoc Query/Filtering Options
- Feature Property Filtering Options
- Array Order Designation/Filtering
- Array Group Assignment/Filtering
- Filter/Group by Array Properties
- Average Arrays within Groups
- Two or more Group Comparison
- PAM: Prediction Analysis for Microarrays
- Boolean Comparison with another Set
- Clustering: Hierarchical
- Clustering: Kmeans
- Clustering: SOM
- Correlation Summary Report
- Gene Ontology Summary Report
- Pathways Summary Report
- Save As a New Dataset**
- Additional Filtering Options

genes.  
genes.  
(4) arrays.  
st 80% (4) array(s)

and [Proceed](#)

At any time, can  
save a subset as a  
new dataset  
In effect, this starts  
the tree of subsets  
over again at the  
top...

## Managing Temporary Datasets

**Access** datasets located in your: [Temporary](#) or [Permanent](#) area

Switch to **managing** datasets located in your: [Permanent](#) area

[Need Help?](#) 

Check boxes to select datasets for action

Temporary Datasets		Created	Containing Arrays	Genes	Gene Information Refreshed
<input checked="" type="checkbox"/>	hands-on qual filter	Dec 12 11:37:02am	5	5276	Dec 12 11:38:27am
Select an Action to perform on selected datasets			Continue		
Select an Action to perform on selected datasets					
Delete the selected datasets			ad Status		
Move the selected datasets to your Permanent Area			ads   GeneCards		

### Dataset Management:

1. Can delete a dataset – but must delete parent and all children!
2. Can promote datasets (Transient to Temporary or Permanent; Temporary to Permanent)

## Interactive Array Filtering

**Arrays Included**

Change Array order. ↑ ↓

Mm-Incyte-v1p1-1 Sample 1/Type A  
Mm-Incyte-v1p1-2 Sample 2/Type A  
Mm-Incyte-v1p1-3 Sample 3/Type A  
Mm-Incyte-v1p1-4 Sample 4/Type A  
Mm-Incyte-v1p1-5 Sample 5/Type A  
Mm-Incyte-v1p1-6 Sample 1/Type B  
Mm-Incyte-v1p1-7 Sample 2/Type B  
Mm-Incyte-v1p1-8 Sample 3/Type B  
Mm-Incyte-v1p1-9 Sample 4/Type B

↓ Remove or Add Back Arrays ↑

Mm-Incyte-v1p1-10 Sample 5/Type B

**Arrays Excluded**

Subset Label:   
(Optional)

**Change Array Order** by highlighting an array name and using the change array order up and down arrows.

**Remove/Add Arrays** by highlighting an array name and using the remove or add arrows  
Enter a label in the **Subset Label** field to have it attached to the resultant subset

Click the **Filter** button when finished or the **Cancel** button to return to the Data Display.

Allows re-ordering and removal of arrays from a subset


Accessed from main pull down list on Data Display Page


# Exporting Data to Other Microarray Analysis Tools


- BRB Array tools export by well ID or by UniGene ID
- GeneSpring export
- MA Explorer export

**Extraction for BRBArrayTools**

















---

**Data Format/Alignment Options** 

Data Alignment : NCI/BRB's BRBArraytools: Separate Files - Alignment by WellID 

**Array Selection** 


- A Submit

	A	mAddbID: Array Name & Short Description
		28733: Mm-Incyte-v1p1-1 Sample 1/Type A
		28742: Mm-Incyte-v1p1-10 Sample 5/Type B
		28734: Mm-Incyte-v1p1-2 Sample 2/Type A
		28735: Mm-Incyte-v1p1-3 Sample 3/Type A
		28736: Mm-Incyte-v1p1-4 Sample 4/Type A
		28737: Mm-Incyte-v1p1-5 Sample 5/Type A
		28738: Mm-Incyte-v1p1-6 Sample 1/Type B
		28739: Mm-Incyte-v1p1-7 Sample 2/Type B



# Retrieving Uploaded Data

## mAddb: Data Retrieval Form

 This tool allows you to retrieve the original uploaded data files.

### Upload Retrieval Options

Package Format:

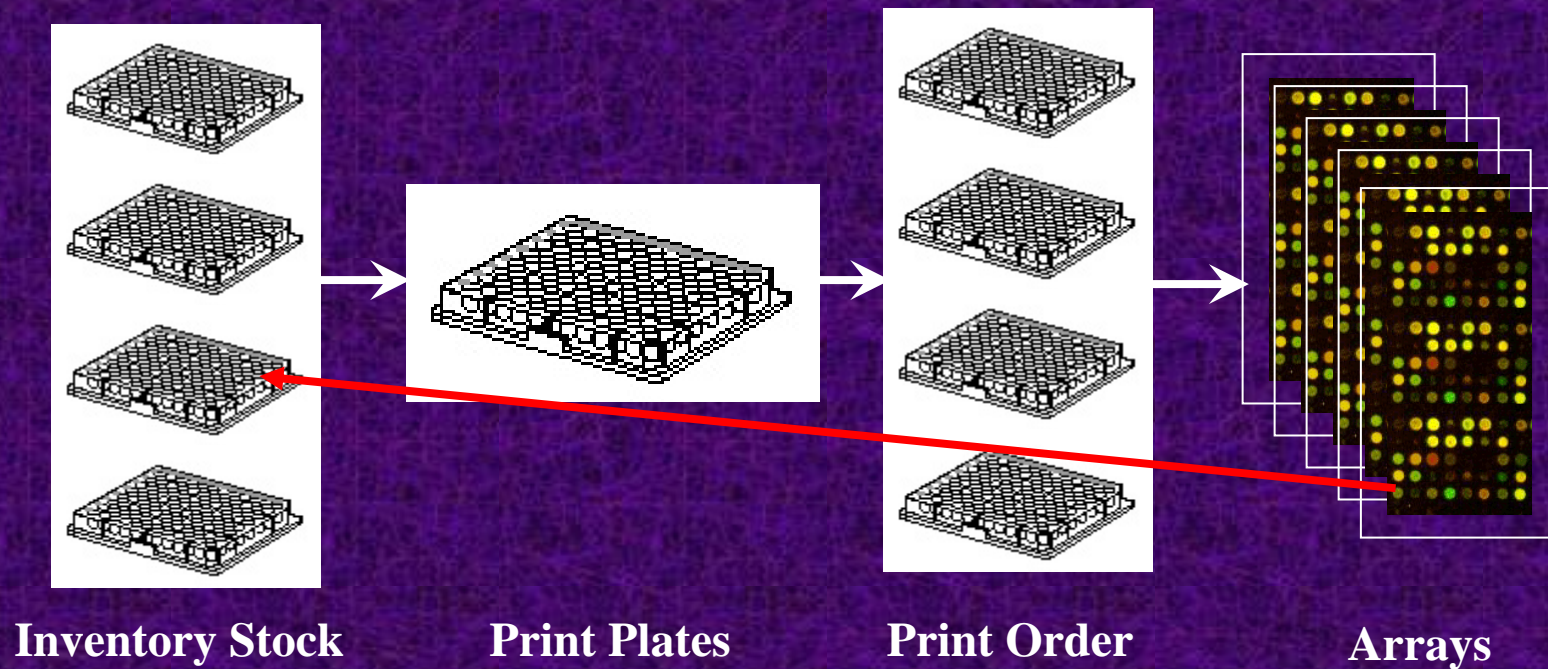
Include: ☒ Image Files (Spotted Uploads only)

☒ Array Description Files

### Array Selection

<input type="button" value="-"/>	<input type="button" value="A"/>	<input type="button" value="Submit"/>		
	A	ID #	Array Name & Description	
<input type="radio"/>	<input checked="" type="radio"/>	28733	Mm-Incyte-v1p1-1 Sample 1/Type A	
<input type="radio"/>	<input checked="" type="radio"/>	28734	Mm-Incyte-v1p1-2 Sample 2/Type A	
<input type="radio"/>	<input checked="" type="radio"/>	28735	Mm-Incyte-v1p1-3 Sample 3/Type A	
<input type="radio"/>	<input checked="" type="radio"/>	28736	Mm-Incyte-v1p1-4 Sample 4/Type A	
<input type="radio"/>	<input checked="" type="radio"/>	28737	Mm-Incyte-v1p1-5 Sample 5/Type A	

# *mAdb* Database Design: Feature Tracking



- mAdb works with microarray facilities to track printing from arrays back to inventory plates
- Allows mAdb support staff to correct printing errors

# mAdb Tips for array analysis

- Always look at Project Summaries – normalization factor for a “good” array should be between 0.5 and 2.0.
- If you have replicate arrays (and you should), do a scatter plot to determine the correlation between the arrays (i.e. how close the slope is to 1. For reverse fluors, how close to  $-1$ ) just for QC purposes.
- Turning **Show Spot Images** off, generally displays results faster – only need for spot QA.



# General tips for array analysis

At a recent Microarray Data Analysis conference in Washington D.C., several speakers laid out what distinguishes a good microarray experiment from a bad one:

- When possible, consult a statistician before you even design your experiment - they offer more than just analysis tools.
- Do a power analysis to determine the number of replicates (i.e. chips) you need to detect an effect. To estimate the effect size, you might want to run a pilot study first or obtain the estimate from previous similar experiments. Regardless of the power analysis results, obtain at least three replicates on different slides or chips.
- Find sources of technical variation before you embark on a hunt for biological effects and standardize your protocols.
- Randomize your variables: for example, don't run all your treatment slides on one day and all your controls on the next.
- Microarray analysis is still a screening tool – confirm your observation by other methods – RT-PCR, Northern blot, protein levels
- See <http://linus.nci.nih.gov/~brb/TechReport.htm> for good references on design, analysis issues, and myths/truths



# Other microarray training:

- Hands-on analysis tool mAdb class #412 – next available class on November 16-17, 1 - 4 PM
- Statistical Analysis of Microarray Data & BRB Array Tools (from the NCI Biometrics Research Branch) class #410 offered bimonthly; next class October 26-27, 12:30-4:30PM
- Partek Pro, R, GeneSpring classes – [training.cit.nih.gov](http://training.cit.nih.gov)
- Sample datasets to try out the system are available from a link on the Gateway Page



## Uploading Links

- [Upload](#) Array data
- [Status](#) of Uploads
- [Upload](#) Identifier lists

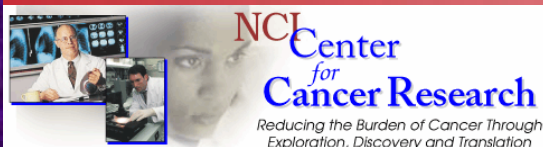


[Access](#) Training/Public Datasets

## mAdb Development and Support Team:

- John Powell, Chief, BIMAS, CIT
- Esther Asaki\*
- John Greene, Ph.D.\*
- Liming Yang, Ph.D
- Kathleen Meyer\*
- Jim Tomlin
- Tim Ruppert\*

\*SRA International contractor



<http://madb.nci.nih.gov>  
<http://madb.niaid.nih.gov>

**For assistance, remember:**

**[madb\\_support@bimas.cit.nih.gov](mailto:madb_support@bimas.cit.nih.gov)**

**Thank you!!**

